HARMONIZED INTEGRATED HAZARD CLASSIFICATION SYSTEM FOR HUMAN HEALTH AND ENVIRONMENTAL EFFECTS OF CHEMICAL SUBSTANCES

As endorsed by the 28th Joint Meeting of the Chemicals Committee and the Working Party on Chemicals in November 1998

TABLE OF CONTENTS

PART I:

GENERAL INTRODUCTION TO THE HARMONIZED INTEGRATED HAZARD CLASSIFICATION SYSTEM

INTRODUCTION

1. The production and use of chemicals is fundamental in the economic development of all countries and, at the same time, it may pose a risk to the health and well-being of all people and the environment if not managed in a responsible manner. The primary objective of hazard classification and communication systems is to provide information to protect human health and the environment. One essential step leading to the safe use of chemicals is the identification of the specific hazards and the organisation of that information so that it can be conveyed to users of chemicals in a form that is easy to understand. Measures can then be taken to avoid or manage potential risks in circumstances where exposure may occur. This is the fundamental rationale behind the hazard classification and labelling of chemicals. It has traditionally led at the national level to sector-specific regulations (transport, industry, environment, health, agriculture, consumer products, occupational health). Because of differences in use and exposure, hazard classification systems usually vary between sectors. In some cases, there is little or no consistency within sectors between different countries.

2. In 1952, the International Labor Office (ILO) began a study of the classification and labelling of dangerous substances which led in 1989 to a Resolution considering the harmonization of systems of classification and labelling for the use of hazardous chemicals at work.

3. In 1953, the UN Economic and Social Council created the UN Committee of Experts on the Transport of Dangerous Goods (UNCETDG) charged with developing recommendations addressed to governments and international organisations concerned with the regulation of the transportation of dangerous goods; amongst other aspects, these recommendations cover the principles of classification and definitions of the classes of dangerous goods. In 1956, the UNCETDG first published its UN Recommendations on Transport of Dangerous Goods (UNRTDG) which were recently modified (1997) for the tenth time. The UNRTDG are now included in the transport legislation of many UN states and they are used by the International Maritime Organisation (IMO), the International Civil Aviation Organisation (ICAO) and other international bodies covering transport modes. Thus land-sea-air transport is the only sector where harmonization of hazard classification and labelling has been to a large degree achieved.

4. The UN Conference on Environment and Development (UNCED) in 1992 identified the harmonization of classification and labelling of chemicals as one of six action programs in Chapter IX of UNCED Agenda 21. Its objective was: “a globally harmonized hazard classification and compatible labelling system (GHS) including material safety data sheets and easily understandable symbols, should be available, if feasible, by the year 2000.” It was recognised that, while a harmonized classification system might be feasible, harmonized labelling may or may not be appropriate or possible across all sectors, but that compatibility of labelling systems might be achievable.

5. UNCED identified the International Program on Chemical Safety (IPCS) as the nucleus for international co-operation on Chapter XIX activities. Under the umbrella of IPCS a Co-ordinating Group for the Harmonization of Chemical Classification Systems (CG/HCCS) was established to promote and oversee the work to develop a GHS. Later, the oversight of the work of the CG/HCCS was provided by the broader Inter Organisational Programme for the Sound Management of Chemicals - IOMC. As expressed in the CG/HCCS Terms of Reference, the goals of international harmonization are to:

- enhance the protection of mankind and the environment by providing an internationally comprehensible system for hazard communication;
- provide a recognised framework for those countries without an existing system;
- reduce the need for testing and evaluation of chemicals;

- facilitate international trade in chemicals whose hazards have been properly assessed and identified on an international basis.

ORGANISATIONAL CONTEXT FOR DEVELOPMENT OF THE GHS

6. The first priority of the CG/HCCS was the development of a harmonized classification system defining the hazards of various endpoints of concern. The Organisation for Economic Co-operation and Development (OECD) was identified as the Focal Point for work on human health and environmental hazards, ILO/UNCETDG as the Focal Point for work on physical hazards, and ILO as the Focal Point for work on Hazard Communication. The CG/HCCS would integrate the harmonized classification scheme with a harmonized hazard communication system to give an overall Globally Harmonized Classification and labelling System (GHS).

The OECD Advisory Group on Harmonization of Classification and Labelling (AG-HCL)

7. The AG-HCL was formally established in 1994 by the Joint Meeting of the OECD Chemicals Group and Management Committee to develop proposals for a harmonized classification system for the hazards of chemicals to human health and the environment. It based its work on the initial efforts of an OECD Clearing House (1991-1993) on the Acute Human Toxicity and on the Acute Aquatic Toxicity of chemicals.

8. In its work the AG-HCL followed a set of general principles developed by the IOMC-GG/HCCS for the work on harmonization of the hazard classification of chemicals, that specifically:.

a) the level of protection offered to workers, consumers, the general public and the environment should not be reduced as a result of harmonizing the classification and labelling systems;

b) the hazard classification process refers only to the hazards arising from the intrinsic properties of chemical elements and compounds, and mixtures thereof, whether natural or synthetic;

c) harmonization means establishing a common and coherent basis for chemical hazard classification and communication, from which the appropriate elements relevant to means of transport, consumer, worker and environment protection can be selected;

d) the scope of harmonization includes both hazard classification criteria and hazard communication tools, e.g. labelling and chemical safety data sheets;

e) changes in all existing systems will be required to achieve a single globally harmonized system; transitional measures should be included in the process of moving to the new system.

f) the involvement of concerned international organisations of employers, workers, consumers, and other relevant organisations in the process of harmonization should be ensured,

g) the comprehension of chemical hazard information, by the target audience, e.g. workers, consumers and the general public, should be addressed;

h) test data already generated for the classification of chemicals under the existing systems, should be accepted when reclassifying these chemicals under the harmonized system;

i) a new harmonized classification system may require adaptation of existing methods for testing of chemicals;

j) in relation to chemical hazard communication, the safety and health of workers, consumers and the public in general should be ensured while protecting confidential business information, as prescribed by the competent authorities.

9. The work of the AG-HCL was generally of three related kinds:

a) Comparison of the major classification systems, identification of similar or identical elements and, for the elements which were dissimilar, development of a consensus on a compromise;

b) Examination of the scientific basis for the criteria which define the end-point of concern, gaining expert consensus on the test methods, data interpretation and level of concern, and then seeking consensus on the criteria. For some end-points, the existing schemes had no criteria and the relevant criteria were developed by the AG-HCL;

c) Where there was a decision-tree approach (e.g. irritation) or where there were dependent criteria in the classification scheme (acute aquatic toxicity), development of consensus on the process or the scheme for using the criteria.

10. The AG-HCL proceeded stepwise in developing its harmonized classification criteria. For each end-point the following steps were undertaken:

Step 1:

A thorough analysis of existing classification systems, including the scientific basis for the system and its criteria, its rationale and explanation of the mode of use. A Step 1 document is prepared and amended as required after discussion by AG-HCL

Step 2:

A proposal for a harmonized classification system and criteria for each class is developed. A Step 2 document is prepared and amended as required after discussion by AG-HCL

Step 3:

(a) AG-HCL reaches consensus on the revised Step 2 proposal; or

(b) After attempts at consensus building fail, the specific non-consensus items are identified as alternatives in a revised Step 2 proposal.

Step 4:

Final proposal is submitted to the OECD Joint Meeting for approval and subsequently to the IOMC CG-HCCS for global implementation.

11. As experience with the use of the system is accumulated, and as new scientific information emerges, the test methods, the interpretation of the test data and the harmonized criteria *per se* may have to be updated. Thus, international work will continue to be needed in the future and, depending on the nature of the future international instrument for the implementation of the GHS, decisions will have to be made on the mechanism for carrying out the updating work in the future.

GENERAL CONSIDERATIONS

Scope of the Harmonized Classification System

12. The work on harmonization of hazard classification and labelling focuses on a harmonized system for all chemicals and mixtures of chemicals. The application of the components of the system may vary by type of product or stage of the life cycle.

13. The classification system applies to pure chemical substances, their dilute solutions and to mixtures of chemical substances. However, since special considerations are needed to classify mixtures, an OECD Working Group on Classification Criteria for Mixtures has begun its work to address harmonization in this area.

15. One objective of the harmonized hazard classification system is for it to be simple and transparent with a clear distinction between classes in order to allow for "self classification" as far as possible. For many end-points the criteria are semi-quantitative or qualitative and expert judgement is required to interpret the data for classification purposes. Furthermore, for some end-points, e.g. eye irritation, a decision tree approach is given as an example.

Presentation of Criteria

16. The current criteria for specific endpoints are presented as a series of chapters in this paper. These chapters include a number of sections all of which are relevant to classification decisions. Some chapters also have an Appendix which, unless clearly indicated to the contrary, are not part of the criteria and should be regarded as background information only. For one endpoint (hazardous for the aquatic environment) a separate Guidance Document is considered essential for a good understanding and use of the system.

Test Methods and Test Data Quality

17. The classification of a chemical substance depends both on the criteria and on the reliability of the test methods underpinning the criteria. In some cases the classification is determined by a pass or fail of a specific test, e.g. the ready biodegradation test, while in other cases, interpretations are made from dose/response curves and observations during testing. In all cases, the test conditions need to be standardized so that the results are reproducible with a given chemical substance and the standardized test yields "valid" data for defining the end-point of concern. In this context, validation is the process by which the reliability and the relevance of a procedure are established for a particular purpose.

18. Tests that determine hazardous properties which are conducted according to internationally recognised scientific principles can be used for purposes of a hazard determination for health and environmental hazards. The GHS criteria for determining health and environmental hazards should be test method neutral, allowing different approaches as long as they are scientifically sound and validated

according to international procedures and criteria already referred to in existing systems for the endpoint of concern and produce mutually acceptable data.

Previously Classified Chemicals

19. One of the general principles established by the IOMC-CG-HCCS states that test data already generated for the classification of chemicals under the existing systems should be accepted when classifying these chemicals under the harmonized system thereby avoiding duplicative testing and the unnecessary use of test animals. This policy has important implications in those cases where the criteria in the GHS are different from those in an existing system. In some cases, it may be difficult to determine the quality of existing data from older studies. In such cases, expert judgement will needed.

Substances Posing Special Problems

20. The effect of a substance on biological and environmental systems is influenced, inter alia, by the physico chemical properties of the substance and the way in which it is biologically available. Some groups of substances present special problems in this respect, for example some polymers and metals.

Animal Welfare

21. The welfare of experimental animals is a concern . This ethical concern includes not only the alleviation of stress and suffering but also, in some countries, the use and consumption *per se* of test animals. Where possible and appropriate, tests and experiments that do not require the use of live animals are preferred to those using sentient live experimental animals. To that end, for certain end-points (skin and eye irritation/corrosion) testing schemes starting with non-animal observation/measurements are included as part of the classification system. For other endpoints such as acute toxicity, alternative animal tests, using fewer animals or causing less suffering are internationally accepted and should be preferred to the conventional LD50 test.

Evidence From Humans

22. For classification purposes, reliable epidemiological data and experience on the effects of chemicals on humans (e.g. occupational data, data from accident data bases) should be taken into account in the evaluation of human health hazards of a chemical. Testing on humans solely for hazard identification purposes is generally not acceptable.

Weight of Evidence

23. For some hazard endpoints, classification results directly when the data satisfy the criteria. For others, classification of a chemical is made on the basis of the total weight of evidence. This means that all available information bearing on the determination of toxicity is considered together, including the results of valid in vitro tests, relevant animal data, and human experience such as epidemiological and clinical studies and well-documented case reports and observations.

24. The quality and consistency of the data are important. Evaluation of substances related to the material under study should be included, as should site of action and mechanism or mode of action study results. Both positive and negative results are assembled together in a single weight of evidence determination.

25. Positive effects which are consistent with the criteria for classification in each chapter, whether seen in humans or animals, will normally justify classification. Where evidence is available from both sources and there is a conflict between the findings, the quality and reliability of the evidence from both sources must be assessed in order to resolve the question for classification. Generally, data of good quality and reliability in humans will have precedence over other data. However, even well-designed and conducted epidemiological studies may lack sufficient numbers of subjects to detect relatively rare but still significant effects, or to assess potentially confounding factors. Positive results from well-conducted animal studies are not necessarily negated by the lack of positive human experience but require an assessment of the robustness and quality of both the human and animal data relative to the expected frequency of occurrence of effects and the impact of potentially confounding factors.

26. Route of exposure, mechanistic information and metabolism studies are pertinent to determining the relevance of an effect in humans. When such information raises doubt about relevance in humans, a lower classification may be warranted. When it is clear that the mechanism or mode of action is not relevant to humans, the substance should not be classified.

27. Both positive and negative results are assembled together in the weight of evidence determination. However, a single positive study performed according to good scientific principles and with statistically and biologically significant positive results may justify classification.

BUILDING BLOCK APPROACH

28. At various times during the development of harmonized classification criteria, concerns have arisen concerning the way a harmonized classification system might be used and whether it would meet the needs of its various end-users.

29. One of the consequences of the application of the classification system is expressed in the IOMC CG/HCCS General Principle (c):

> "harmonization means establishing a common and coherent basis for chemical hazard classification and communication, from which the appropriate elements relevant to means of transport, consumer, worker and environment protection can be selected."

30. In the following chapters, sufficient sub-classes have been included under some endpoints to accommodate the fundamental needs of the existing systems. The application of the classification scheme may vary according to the circumstances, type of product and stage of the life cycle of the chemical.

31. It is essential that the cut-offs be recognised as a fundamental basis for the harmonized classification system. The use of different cut-offs for any use of the classification system would be contrary to harmonization.

PART 2:

THE HARMONIZED INTEGRATED HAZARD CLASSIFICATION SYSTEM

HARMONIZED SYSTEM FOR THE CLASSIFICATION OF CHEMICALS WHICH CAUSE ACUTE TOXICITY

PURPOSE, BASIS AND APPLICABILITY

1. The purpose of this document is to present a harmonized system of classification for acute toxicity by the oral, dermal, and inhalation routes to be used internationally.

2. The basis for the harmonized criteria are those which are currently in use in OECD countries as well as those recommended by the United National Committee of Experts on the Transport of Dangerous Goods (UNCETDG). Elements from these sources have been integrated so as a to maintain a high level of protection under a globally harmonised system of classification.

3. The classification scheme included elements that will be used by all authorities as well as other categories that will be applied only by some (e.g. transport).

CLASSIFICATION CATEGORIES

4. Chemicals can be allocated to one of five toxicity classes based on acute toxicity by the oral, dermal or inhalation route according to the numeric criteria expressed as (approximate) LD50 (oral, dermal) or LC50 (inhalation) values are shown in the table below. Explanatory notes are shown in italics following the table.

	Class 1	Class 2	Class 3	Class 4	Class 5
Oral (mg/kg)	5	50	300	2000	5000 See detailed criteria
Dermal mg/kg)	50	200	1000	2000	
Gases (ppm) Note a	100	500	2500	5000	
Vapours (mg/l) Note a Note b Note c	0.5	2.0	10	20	
Dusts and Mists (mg/l) Note a Note d	0.05	0.5	1.0	5	

Notes:

a: *Inhalation cut-off values in the table are based on 4 hour testing exposures. Conversion of existing inhalation toxicity data which has been generated according to 1 hour exposures should be by dividing by a factor of 2 for gases and vapours and 4 for dusts and mists.*

b: *It is recognised that saturated vapour concentration may be used as an additional element by some regulatory systems to provide for specific health and safety protection. (e.g. UN Recommendations for the Transport of Dangerous Goods).*

c: *For some chemicals the test atmosphere will not just be a vapour but will consist of a mixture of liquid and vapour phases. For other chemicals the test atmosphere may consist of a vapour which is near the gaseous phase. In these latter cases, classification should be based on ppm as follows: Class 1 (100 ppm), Class 2 (500 ppm), Class 3 (2500 ppm), Class 4 (5000 ppm). Work in the OECD Test Guidelines Programme should be undertaken to better define the terms "dusts", "mists" and "vapours" in relation to inhalation toxicity testing.*

d: *The values for dusts and mists should be reviewed to adapt to any future changes to OECD Test Guidelines with respect to technical limitation in generating, maintaining and measuring dust and mist concentrations in respirable form.*

CRITERIA FOR CLASS 5

5. Criteria for class 5 are intended to enable the identification of substances which are of relatively low acute toxicity hazard but which, under certain circumstances may present a danger to vulnerable populations. These substances are anticipated to have an oral or dermal LD50 in the range of 2000-5000 mg/kg or equivalent doses for other routes.

6. The specific criteria for class 5 are :

a) The substance is classified in this Class if reliable evidence is already available that indicates the LD50 or (LC50) to be in the range of class 5 values or other animal studies or toxic effects in humans indicate a concern for human health or an acute nature.

b) The substance is classified in this Class, through extrapolation, estimation or measurement of data, if assignment to a more hazardous class is not warranted, and :

- reliable information is available indicating significant toxic effects in humans; or
- any mortality is observed when tested up to Class 4 values by the oral, inhalation, or dermal routes; or
- where expert judgement confirms significant clinical signs of toxicity, when tested up to Class 4 values, except for diarrhoea, piloerection or an ungroomed appearance, or
- where expert judgement confirms reliable information indicating the potential for significant acute effects from other animal studies.

7. Recognising the need to protect animal welfare, testing in animals in class 5 ranges is discouraged and should only be considered when there is a strong likelihood that results of such a test would have a direct relevance for protecting human health.

RATIONALE FOR THE PROPOSED SYSTEM

General considerations

8. The harmonized classification system for acute toxicity has been developed in such a way as to accommodate the needs of existing systems. A basic principle set by the IOMC CG/HCCS is that "harmonization means establishing a common and coherent basis for chemical hazard classification and communication from which the appropriate elements relevant to means of transport, consumer, worker and environment protection can be selected." To that end, five classes have been included in the acute toxicity scheme.

9. The preferred test species for evaluation of acute toxicity by the oral and inhalation routes is the rat, while the rat or rabbit are preferred for evaluation of acute dermal toxicity. As noted by the CG/HCCS, "Test data already generated for the classification of chemicals under existing systems should be accepted when reclassifying these chemicals under the harmonized system." When experimental data for acute toxicity are available in several animal species, scientific judgement should be used in selecting the most appropriate LD50 value from among valid, well-performed tests.

10. Class 1, the highest toxicity class, has cut off values of 5 mg/kg by the oral route, 50 mg/kg by the dermal route, 100 ppm for gases or gaseous vapours, 0.5 mg/l for vapours, and 0.05 mg/l for dusts and mists. These toxicity values are currently used primarily by the transport sector for classification for packing groups.

11. Class 5 is for chemicals which are of relatively low acute toxicity but which, under certain circumstances, may pose a hazard to especially vulnerable populations. Criteria for identifying substances in class 5 are provided in addition to the table. These substances are anticipated to have an oral or dermal LD50 value in the range 2000 - 5000 mg/kg or equivalent doses for other routes of exposure.. In light of animal welfare considerations, testing in animals in class 5 ranges is discouraged and should only be considered when there is a strong likelihood that results of such testing would have a direct relevance for protecting human health.

Special considerations for inhalation toxicity

12. Values for inhalation toxicity are based on 4 hour tests in laboratory animals. When experimental values are taken from tests using a 1 hour exposure, they can be converted to a 4 hour equivalent by dividing the 1 hour value by a factor of 2 for gases and vapours and 4 for dusts and mists.

13. Units for inhalation toxicity are a function of the form of the inhaled material. Values for dusts and mists are expressed in mg/l. Values for gases are expressed in ppm. Acknowledging the difficulties in testing vapours, some of which consist of mixtures of liquid and vapours phases, the table provides values in units of mg/l. However, for those vapours which are near the gaseous phase, classification should be based on ppm. As inhalation test methods are updated, the OECD and other test guideline programs will need to define vapours in relation to mists for greater clarity.

14. Vapour inhalation values are intended for use in classification of acute hazard for all sectors. It is also recognised that the saturated vapour concentration of a chemical is used by the transport sector as an additional element in classifying chemicals for packing groups.

15. Of particular importance is the use of well articulated values in the high toxicity classes for dusts and mists. Inhaled particles between 1 and 4 microns mean mass aerodynamic diameter (MMAD) will deposit in all regions of the rat respiratory tract. This particle size range corresponds to a maximum dose of about 2 mg/l. In order to achieve applicability of animal experiments to human exposure, dusts and mists would ideally be tested in this range in rats. The cut off values in the table for dusts and mists allow clear distinctions to be made for materials with a wide range of toxicities measured under varying test conditions. The values for dusts and mists should be reviewed in the future to adapt to any future changes in OECD or other test guidelines with respect to technical limitations in generating, maintaining, and measuring dust and mist concentrations in respirable form.

HARMONIZED SYSTEM FOR THE CLASSIFICATION OF CHEMICALS WHICH CAUSE SKIN IRRITATION/CORROSION

EXECUTIVE SUMMARY

1. From a comparison of existing dermal irritation/corrosion classification procedures currently in use, a harmonized system was formulated. It includes an evaluation strategy of existing information and specific testing for dermal effects. In developing potential harmonized positions for dermal irritation/corrosion testing, two objectives have been kept in mind: to define criteria for both corrosion and irritation classification that are in the range of sensitivity of existing systems and to have the possibility of subdividing effects into different subclasses for those authorities that need them.

2. A single class is adopted for skin corrosion. Authorities wanting to have up to three subclasses may subdivide the single corrosive class. These subclasses are modelled after those currently in use in the United Nations transport authority.

3. A single class is adopted for skin irritation. The classification procedure draws upon those currently employed by the European Union (EU). Erythema/eschar and edema are graded separately; an animal's mean score from readings over the first three days after exposure must meet a defined level to be positive; and at least 2 of 3 tested animals must be positive for the test to be positive. Positive responses can also be obtained using other, less common criteria. The proportion of test substances expected to be positive by the proposed irritant class is within the range of positives among existing classification systems; it is somewhat higher than that of some of the current classification systems but below those of other systems. Authorities wanting to have two hazard classes can use both irritant and mild irritant classes.

PURPOSE, BASIS AND APPLICABILITY

4. The purpose of the document is to present a harmonized system of classification for skin irritation and corrosion that can be agreed upon and utilised internationally.

5. The harmonized classification system grew out of the major systems that are currently employed. It is based on concepts already in effect and does not deviate significantly from those currently in use.

6. The harmonized system for classification of skin irritation and corrosion include elements that are harmonized and will be used by all authorities as well as other categories that will be applied by only some authorities (e.g., transport, pesticides).

CLASSIFICATION CATEGORIES AND CRITERIA

7. The harmonized system includes guidance for the use of initial considerations, that is those data elements that are evaluated before animal testing for dermal corrosion and irritation is undertaken. It also includes hazard classes for corrosion and irritation.

Initial Considerations

8. Several factors should be considered in determining the corrosion and irritation potential of chemicals before testing is undertaken. Existing human experience and data including from single or repeated exposure and animal observations and data should be the first line of analysis, as it gives information directly referable to effects on the skin. In some cases enough information may be available from structurally related compounds to make classification decisions. Likewise, pH extremes like $\leq$ 2 and $\geq$ 11.5, may indicate dermal effects, especially when buffering capacity is known, although the correlation is not perfect. Generally, such agents are expected to produce significant effects on the skin. It also stands to reason that if a chemical is highly toxic by the dermal route, a dermal irritation/corrosion study may not be practicable since the amount of test substance to be applied would considerably exceed the toxic dose and, consequently, would result in the death of the animals. When observations are made of dermal irritation/corrosion in acute toxicity studies and are observed up through the limit dose, additional testing would not be needed, provided that the dilutions used and species tested are equivalent. In vitro alternatives that have been validated and accepted may also be used to help make classification decisions.

9. All the above information that is available on a chemical should be used in determining the need for in vivo dermal irritation testing. Although information might be gained from the evaluation of single parameters within a tier (e.g., caustic alkalies with extreme pH should be considered as dermal corrosives), there is merit in considering the totality of existing information and making an overall weight of evidence determination. This is especially true when there is information available on some but not all parameters. Generally, primary emphasis should be placed upon existing human experience and data, followed by animal experience and testing data, followed by other sources of information, but case-by-case determinations are necessary.

10. A tiered approach to the evaluation of initial information should be considered, where applicable (Figure 1), recognising that all elements may not be relevant in certain cases.

Corrosion

11. A single harmonized corrosion class is adopted using the results of animal testing. A corrosive is a test material that produces destruction of skin tissue, namely, visible necrosis through the epidermis and into the dermis) in $\geq$ 1 of 3 tested animals after exposure up to a 4 hour duration. Corrosive reactions are typified by ulcers, bleeding, bloody scabs and, by the end of observation at 14 days, by discoloration due to blanching of the skin, complete areas of alopecia and scars. Histopathology should be considered to discern questionable lesions.

Figure 1. TIERED TESTING AND EVALUATION OF DERMAL CORROSION AND IRRITATION POTENTIAL (see also the "Testing and Evaluation Strategy for Eye Irritation/Corrosion")

Step	Parameter	Finding	Conclusion
1a	Existing human or animal experience[g)] ⇒	Corrosive	Classify as corrosive [a]
	⇓ Not corrosive or no data ⇓		
1b	Existing human or animal experience[g)] ⇒	Irritant	Classify as irritant [a]
	⇓ Not irritant or no data ⇓		
1c	Existing human or animal experience ⇒	Not corrosive or irritant	No further testing
	⇓ No data ⇓		
2a	Structure-activity relationships or structure-property relationships [b] ⇒	Corrosive	Classify as corrosive [a]
	⇓ Not corrosive or no data ⇓		
2b	Structure-activity relationships or structure-property relationships [b] ⇒	Irritant	Classify as irritant [a]
	⇓ Not irritating or no data ⇓		
3	pH with buffering [c] ⇒	pH $\leq$ 2 or $\geq$11.5	Classify as corrosive [a]
	⇓ Not pH extreme or no data ⇓		
4	Existing dermal data in animals indicate no need for animal testing [d] ⇒	Yes	Possibly no further testing may be deemed corrosive/irritant
	⇓ No indication or no data ⇓		
5	Valid and accepted in vitro dermal corrosion test [e] ⇒	Positive response	Classify as corrosive [a]
	⇓ Negative response or no data ⇓		

Figure 1. (continued) TIERED TESTING AND EVALUATION OF DERMAL CORROSION AND IRRITATION POTENTIAL

Step	Parameter		Finding	Conclusion
6	Valid and accepted in vitro dermal irritation test [f]	⇒	Positive response	Classify as irritant [a]
	⇓			
	Negative response or no data			
	⇓			
7	In vivo dermal corrosion test (1 animal)	⇒	Corrosive response	Classify as corrosive [a]
	⇓			
	Negative response			
	⇓			
8	In vivo dermal irritation test (3 animals total) [h]	⇒	Irritant response	Classify as irritant
	⇓			
	Negative response	⇒	No further testing	
	⇓			
9	When it is ethical to perform human patch testing [g]	⇒	Irritant response	Classify as irritant [a]
	⇓			
	Not as above	⇒	Nonirritant response	No further testing

a Classify in the harmonized class, below.

b Structure-activity and structure-property relationships are presented separately but would be conducted in parallel.

c Measurement of pH alone may be adequate, but assessment of acid or alkali reserve is preferable; methods are needed to assess buffering capacity.

d Pre-existing animal data should be carefully reviewed to determine if in vivo dermal corrosion/irritation testing is needed. As examples, testing may not be needed when a test material has not produced any dermal irritation in an acute dermal toxicity test at the limit dose, or produces very toxic effects in an acute dermal toxicity test. In the latter case, the material would be classed as being very hazardous by the dermal route for acute toxicity; it is moot whether the material is also irritating or corrosive on the skin. It should be kept in mind in evaluating acute dermal toxicity information that the reporting of dermal lesions may be incomplete, testing and observations may be made on a species other than the rabbit, and species may differ in sensitivity in their responses.

e Currently there are no internationally accepted validated in vitro methods of dermal corrosion, but a validation study on several methods has just been completed.

f Presently there are no validated and internationally accepted in vitro test methods for dermal irritation.

g This evidence could be derived from single or repeated exposures. There is no internationally accepted test method for human dermal irritation testing, but an OECD guideline has been proposed.

h Testing is usually conducted in 3 animals, one coming from the negative corrosion test.

12. For those authorities wanting more than one designation of corrosivity, up to three subclasses are adopted which divide up responses in the corrosive class (Class 1 see Table 1): **subclass 1A** --where responses are noted following up to 3 minutes exposure and up to 1 hour observation; **subclass 1B** --where responses are described following exposure between 3 minutes and 1 hour and observations up to 14 day; and **subclass 1C** --where responses occur after exposures between 1 hour and 4 hours and observations up to 14 days.

Table 1. CORROSIVE CLASS AND SUBCLASSES[a)]

Corrosive class (Class 1)	**Potential corrosive subclasses**	**Corrosive in ≥ 1 of 3 animals**	
(applies to authorities not using subclasses)	(only applies to some authorities)	**exposure**	**observation**
corrosive	corrosive subclass 1A	≤ 3 minutes	≤ 1 hour
	corrosive subclass 1B	> 3 minutes -- ≤ 1 hour	≤ 14 days
	corrosive subclass 1C	> 1 hour -- ≤ 4 hours	≤ 14 days

a. In case human data is considered, the use of human data is discussed under "general considerations", in the introductory chapter of the Harmonized Integrated Classification System.

Irritation

13. A single irritant class is adopted that (a) is centrist in sensitivity among existing classifications, (b) recognises that some test materials may lead to effects which persist throughout the length of the test, and (c) acknowledges that animal responses in a test may be quite variable. The current EU 3-animal classification system is modified to generate the proposed position. An additional mild irritant class is available for those authorities that want to have more than one dermal irritant category.

14. Reversibility of dermal lesions is another consideration in evaluating irritant responses. When inflammation persists to the end of the observation period in 2 or more test animals, taking into consideration alopecia (limited area), hyperkeratosis, hyperplasia and scaling, then a material should be considered to be an irritant.

15. Animal irritant responses within a test can be quite variable, as they are with corrosion. A separate irritant criterion should be added to accommodate cases when there is a significant irritant response but less than the mean score criterion for a positive test. For example, a test material might be designated as an irritant if 1 of 3 tested animals shows a very elevated mean score throughout the study, including lesions persisting at the end of an observation period of normally 14 days. Other responses could also fulfil this criterion. However, the responses should be ascertained as being the result of chemical exposure. Addition of this criterion increases the sensitivity of the classification system beyond that of the current EU system.

16. To counterbalance the increases in sensitivity of a designation of an irritant position and to make room for a mild irritant class, the endpoint mean score for a positive animal response is raised from ≥ 2.0 under the current EU system to ≥ 2.3. From a training set of data, the proportion of positive tests for the total data base decreases from 0.59 for the current EU system to 0.34. The exact proportion of positive test materials in the proposed system is not known, but it would definitely be higher than 0.34 and, thus, closer to the proportion of positives in the current EU system. In addition, the proportion of positives will vary considerably with the composition of materials being tested. From the training set, about 0.34 of the chemicals are in the mild irritant class, and the total is the sum of the proportion of irritants and mild irritants, or 0.68 of the chemicals.

17. A single **irritant** class (Class 2) is adopted using the results of animal testing. Authorities (e.g., pesticides) also have available a less severe **mild irritant** class (Class 3). Several criteria distinguish the two classes (Table 2). They mainly differ in the severity of dermal reactions. The major criterion for the irritant class is that at least 2 tested animals have a mean score of $\geq$ 2.3 - $\leq$ 4.0. For the mild irritant class, the mean score cutoffs are $\geq$ 1.5 - < 2.3 for at least 2 tested animals. Test materials in the irritant class would be excluded from being placed in the mild irritant class.

Table 2. IRRITANT CLASS AND SUBCLASS [a]

Classes	Criteria
Irritant (Class 2) (applies to all authorities)	(1) Mean value of $\geq$ 2.3 - < 4.0 for erythema/eschar or for edema in at least 2 of 3 tested animals from gradings at 24, 48 and 72 hours after patch removal or, if reactions are delayed, from grades on 3 consecutive days after the onset of dermal reactions, or (2) Inflammation that persists to the end of the observation period normally 14 days in at least 2 animals, particularly taking into account alopecia (limited area), hyperkeratosis, hyperplasia, and scaling, or (3) In some cases where there is pronounced variability of response among animals, with very definite positive effects related to chemical exposure in a single animal but less than the criteria above.
Mild Irritant (Class 3) (applies to only some authorities)	Mean value of $\geq$ 1.5 - < 2.3 for erythema/eschar or for edema from gradings in at least 2 of 3 tested animals from grades at 24, 48 and 72 hours or, if reactions are delayed, from grades on 3 consecutive days after the onset of dermal reactions (when not included in the irritant class above).

a. In case human data is considered, the use of human data is discussed under "general considerations", in the General Introduction to the Harmonized Integrated Hazard Classification System.

HARMONIZED SYSTEM FOR THE CLASSIFICATION OF CHEMICALS WHICH CAUSE EYE IRRITATION/CORROSION

EXECUTIVE SUMMARY

1. In the following harmonized system for eye irritation/corrosion hazard classification the collection of test guidelines and classification schemes worked out by the EC, the tier scheme of the U.S. regulators, the experiences of the German regulators based on the EU chemicals notification procedure and the outcome of the "OECD Workshop on Harmonization of Validation Criteria for Alternative Tests / Harmonization and Acceptance Criteria for Alternative Toxicological Test Methods" in Solna, Sweden (22nd -24th January, 1996) have been considered.

2. Also reflected are eye irritation/corrosion classification schemes for chemicals which are in force in the member countries of the Organisation for Economic Co-operation and Development, OECD (6), in the European Economic Community, EU and the Canadian Pest Management Regulatory Agency and the Canadian workplace system, WHMIS. Within the transport sectors of the United Nations, UN, only dermal corrosivity is taken into account; eye corrosivity or eye irritating properties are not included within the "Orange Book" of the UN.

3. The harmonized system includes an evaluation strategy of existing information and specific testing for eye effects. In developing harmonized positions for eye irritation/corrosion testing, three objectives have been kept in mind:

- to define criteria for both serious damage to eyes and eye irritation that are in the range of sensitivity of existing systems,
- to have the option of subdividing effects in two parts for those authorities that need them, and
- to avoid testing for local effects on eyes with skin corrosive substances.

4. A single harmonized hazard group is defined for the classification of serious damage to eyes. Serious damage to eyes is defined as severe irreversible effects on the eye including not only corrosive effects like destruction of cornea or conjunctivae but also persistent indication of serious impairment of sight.

5. A single harmonized hazard group is defined for the classification of eye irritation that reverses within an appropriate observation time. The proposed harmonized classification of reversible eye irritation draws upon procedures currently employed by the European Union (EU) and by regulatory authorities in the United States of America (USA) and in Canada. Classified are local effects detected in a Draize test with rabbits that reverse within 21 days after instillation of the substance into the eye. Effects on the cornea, effects on the iris and conjunctival erythema and edema are graded separately; an animal's mean score from readings over the first three days after instillation must meet a defined level to be positive, and at least 2 of 3 tested animals must be positive for the test to be positive. The proportion of test substances expected to be positive by the proposed harmonized system is somewhat higher than that of the current EU system but less than that of the current US and Canadian systems. Authorities wanting to distinguish between mild and moderate eye irritants have the option to use a subcategorization that considers the differences within the current classification systems.

PURPOSE, BASIS AND APPLICABILITY

6. The purpose of the document is to present a harmonized system of hazard classification for eye irritation, destruction of eye tissues and other serious damage to tissues and function of eyes that can be agreed upon and utilised by OECD Member countries.

7. A tiered testing and evaluation scheme is presented that combines pre-existing information on local corrosivity and on eye irritation (including data relating to historical human or animal experience) as well as considerations on structure-activity relationships (SAR) or structure-property relationships (SPR) and the output of validated *in vitro* tests in order to avoid unnecessary animal testing.

8. The harmonized hazard classification system grew out of the currently employed systems within the OECD Member countries. It is based on concepts already in effect and melds together a position that does not deviate significantly from those currently in use.

9. The proposals for classification of eye irritation and serious damage to the eye include elements that are harmonized and will be used by all authorities as well as optional subcategories that will be applied by only some authorities (e.g., authorities classifying pesticides).

CLASSIFICATION CATEGORIES AND CRITERIA

10. The harmonized system includes guidance for the use of initial considerations, that is those data elements that are evaluated before animal testing for eye damaging effects is undertaken. It also includes hazard classes for local lesions on the eyes.

Initial considerations / tier testing and evaluation strategy

11. Before there is any *in vivo* dermal or eye irritation/corrosion testing all existing information on a test material should be reviewed. Preliminary decisions can often be made from them as to whether an agent is corrosive. If a test material can be classified, no testing is required. A highly recommended way of evaluating existing information on agents or of approaching new uninvestigated substances, is to utilise a tier testing strategy for eye irritation/corrosion.

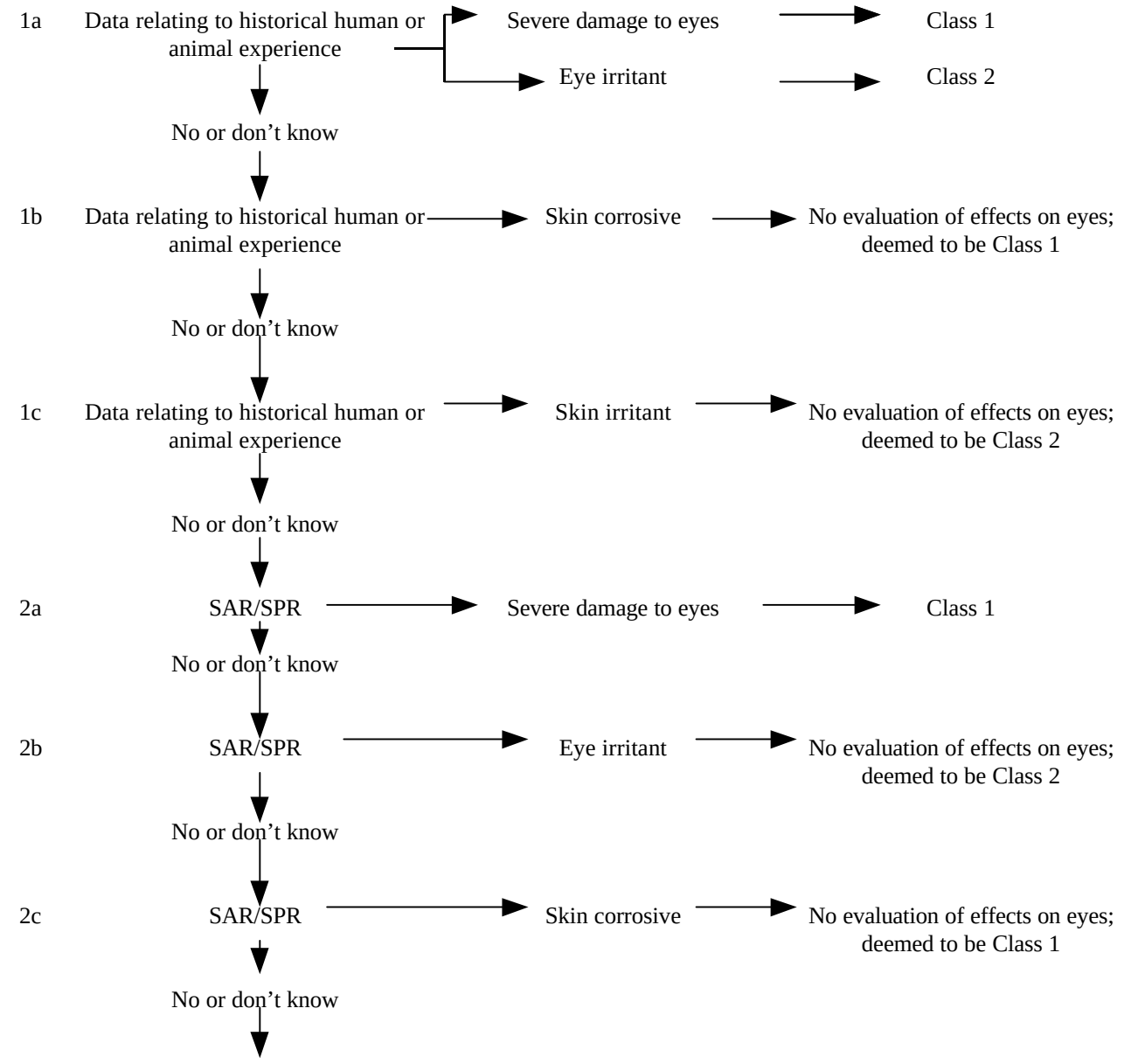
12. Several factors should be considered in determining the eye damage or irritation potential of chemicals before testing is undertaken. Accumulated human and animal experience should be the first line of analysis, as it gives information directly referable to effects on the eye. In some cases enough information may be available from structurally related compounds to make hazard decisions. Likewise, pH extremes like ≤ 2 and ≥ 11.5, may indicate corrosive effects, especially when buffering capacity is known. Such agents are expected to produce significant effects on the eyes. Possible skin corrosion has to be evaluated prior to consideration of eye irritation/corrosion in order to avoid testing for local effects on eyes with skin corrosive substances. *In vitro* alternatives that have been validated and accepted may be used to make classification decisions.

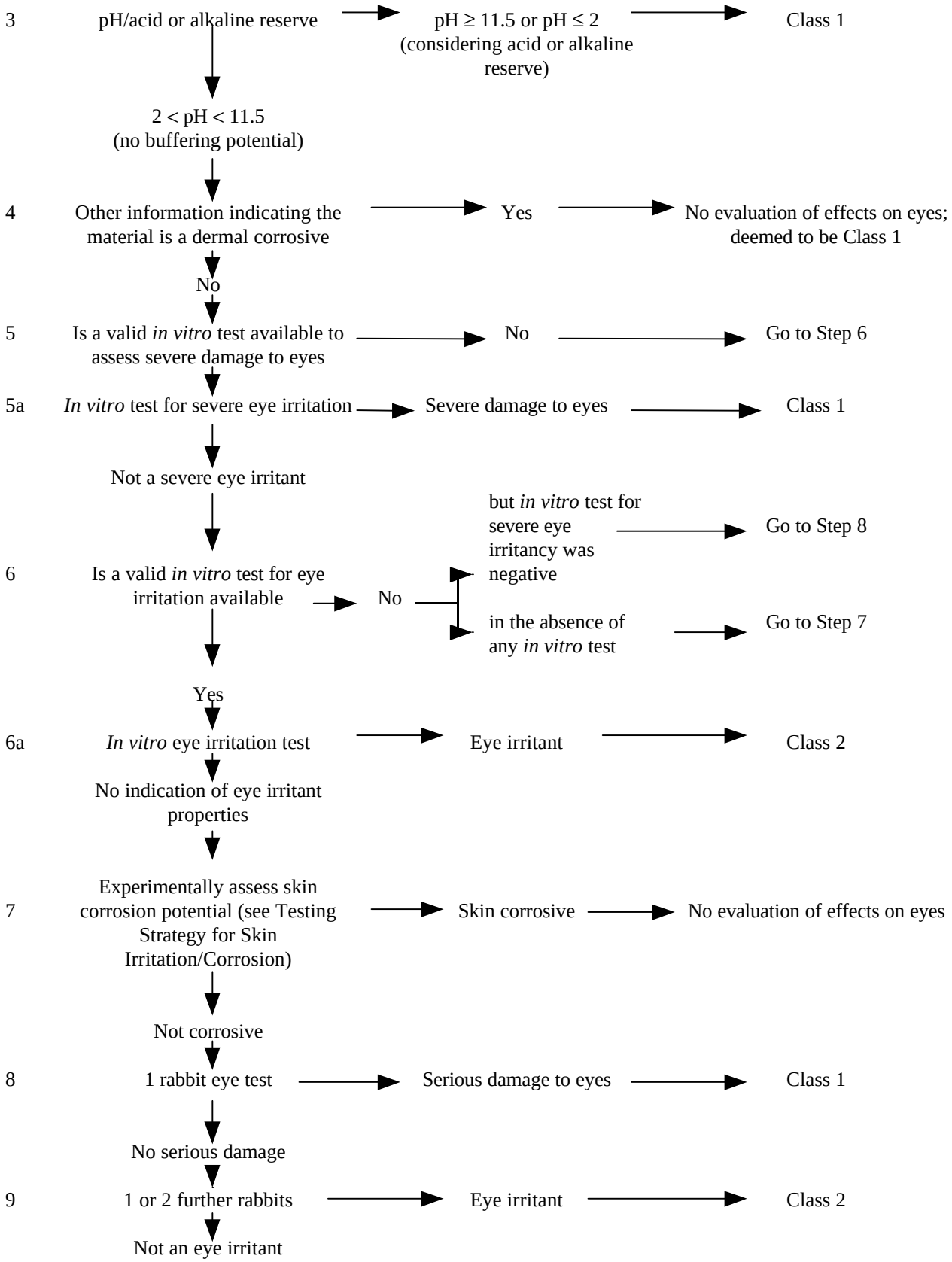
13. All the above information that is available on a chemical should be used in determining the need for *in vivo* eye irritation testing. Although information might be gained from the evaluation of single parameters within a tier (e.g., caustic alkalies with extreme pH should be considered as local corrosives), there is merit in considering the totality of existing information and making an overall weight of evidence determination. This is especially true when there is information available on some but not all parameters. Generally, primary emphasis should be placed upon expert judgement considering human experience with

the substance, followed by the outcome of skin irritation testing and of well validated alternative methods. Animal testing with corrosive substances should be avoided whenever possible.

14. A tiered approach to the evaluation of initial information should be considered, where applicable recognising that all elements may not be relevant in certain cases. The tiered approach explained in figure 1 was developed with contributions from (inter)national centres and committees for the testing and validation of alternatives to animal testing during a workshop in Solna, Sweden.

Figure 1: TESTING AND EVALUATION STRATEGY FOR EYE IRRITATION/CORROSION
(see also the: "Testing and Evaluation Strategy for Skin Irritation/Corrosion")




3
pH/acid or alkaline reserve
pH ≥ 11.5 or pH ≤ 2 (considering acid or alkaline reserve)
Class 1
2 < pH < 11.5 (no buffering potential)
4
Other information indicating the material is a dermal corrosive
Yes
No evaluation of effects on eyes; deemed to be Class 1
No
5
Is a valid *in vitro* test available to assess severe damage to eyes
No
Go to Step 6
5a
In vitro test for severe eye irritation
Severe damage to eyes
Class 1
Not a severe eye irritant
6
Is a valid *in vitro* test for eye irritation available
No
but *in vitro* test for severe eye irritancy was negative
Go to Step 8
in the absence of any *in vitro* test
Go to Step 7
Yes
6a
In vitro eye irritation test
Eye irritant
Class 2
No indication of eye irritant properties
7
Experimentally assess skin corrosion potential (see Testing Strategy for Skin Irritation/Corrosion)
Skin corrosive
No evaluation of effects on eyes
Not corrosive
8
1 rabbit eye test
Serious damage to eyes
Class 1
No serious damage
9
1 or 2 further rabbits
Eye irritant
Class 2
Not an eye irritant

Notes to the testing and evaluation strategy for eye irritation / corrosion

15. Step 1a/b: Data relating to historical human or animal experience: Pre-existing information on eye irritation and skin corrosion are shown separately because evaluation of skin corrosion has to be considered if there is no information on local effects on eyes. Analysis of pre-existing experience with the chemical may identify both corrosion and irritation potential for both dermal and ocular effects: i) Step 1a - reliable determination of eye irritancy basing on human or animal experience - depends on expert judgement: In most cases human experience is based on accidental events and thus, the local effects detected after an accident have to be compared with classification criteria created for evaluation of animal test data. ii) Step 1b - evaluation of data on skin corrosivity - skin corrosive substances should not be instilled into the eyes of animals; such substances should be considered as corrosive to the eyes as well. **(Class 1)**

16. Step 2a/b: SAR (Structure Activity Relationships) / SPR (Structure Property Relationships) for eye irritation and skin corrosion are shown separately but in reality would probably be done in parallel. This stage should be completed using validated and accepted SAR/SPR approaches. The SAR/SPR analysis may identify both corrosion and irritation potential for both dermal and ocular effects: i) Step 2a - reliable determination of eye irritancy only by theoretical evaluations - in most cases it will only be appropriate for substances that are homologous to agents with very well known properties. ii) Step 2c - theoretical evaluation of skin corrosivity - skin corrosive substances should not be instilled into the eyes of animals; such substances should be considered as corrosive to the eyes as well. **(Class 1)**

17. Step 3: pH extremes like <2 and >11.5 may indicate strong local effects, especially in combination with assessment of acid or alkaline reserve (see annexed draft of a respective guideline), substances exhibiting such physico-chemical properties should be considered as corrosive to eyes. **(Class 1)**

18. Step 4: All attainable information should be used, including probable human experience. But this information should be restricted to that which pre-exists (e.g. the results of a dermal LD50 test or historical information on dermal corrosion).

19. Step 5: These must be alternative methods for the assessment of severe eye irritation/corrosion or serious damage to eyes (e.g., irreversible corneal opacity) which have been validated in accordance with internationally agreed principles and criteria (see “General Considerations” of the General Introduction to the Harmonized Integrated Hazard Classification System).

20. Step 6: At present this step seems not be achievable in the near future. Validated alternative methods for the reliable assessment of (reversible) eye irritation need to be worked out.

21. Step 7: In the absence of any other relevant information, it is essential to obtain this via an internationally recognised corrosion/irritation test before proceeding to a rabbit eye irritation test. This must be conducted in a staged manner. If possible, this should be achieved using a validated, accepted in vitro skin corrosivity assay. If this is not available, then the assessment should be completed using animal tests (see the skin irritation/corrosion strategy).

22. Step 8: Staged assessment of eye irritation in vivo. If in a limit test with one rabbit serious damage to eyes/severe eye irritation/corrosion is detected no further testing is needed.

23. Step 9: Only two animals may be employed for irritation testing (including the one used for evaluation of possible severe effects) if these two animals give concordant clearly irritant or clearly non-

irritant responses. In the case of different or borderline responses a third animal is needed. Depending on the result of this three-animal test, classification may be required or not.

24. Where data needed for such a testing strategy cannot be required, the proposed tier testing approach demonstrates a good guidance how to organise existing information on a test material and to make a weight-of-evidence decision about hazard assessment and hazard classification - ideally without conducting new animal tests.

Irreversible effects on the eye / serious damage to eyes

25. A single harmonized hazard class is adopted for substances that have the potential to damage the eyes seriously. This hazard class - Class 1(irreversible effects on the eye) - includes the criteria listed below. These observations include animals with grade 4 cornea lesions and other severe reactions (e.g., destruction of cornea) observed at any time during the test, as well as persistent corneal opacity, discoloration of the cornea by a dye substance, adhesion, pannus, and interference with the function of the iris or other effects that impair sight. In this context, persistent lesions are considered those which are not fully reversible within an observation period of normally 21 days. Hazard classification: Class 1 also contains substances fulfilling the criteria of corneal opacity ≥ 3 or iritis > 1.5 detected in a Draize eye test with rabbits, because severe lesions like these usually do not reverse within a 21 days observation period.

IRREVERSIBLE EFFECTS CLASS

An eye irritant Class 1 (irreversible effects on the eye) is a test material that produces:

- at least in one animal effects on the cornea, iris or conjunctiva that are not expected to reverse or have not fully reversed within an observation period of normally 21 days

  and/or

- at least in 2 of 3 tested animals a positive response of:

  corneal opacity ≥ 3 and/or
  iritis > 1.5

  calculated as the mean scores following grading at 24, 48 and 72 hours after installation of the test material.

26. The use of human data is discussed under "General Considerations" in the introductory chapters of the Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemicals.

Reversible effects on the eye

27. A single category is adopted for substances that have the potential to induce reversible eye irritation. This single hazard category provides the option to identify within the category a sub-category for substances inducing eye irritant effects reversing within an observation time of 7 days.

28. Those authorities desiring one single category for classification of "eye irritation" may use the overall harmonized Class 2 (irritating to eyes): others may want to distinguish between Class 2 (irritating to the eyes) and Class 2A (mildly irritating to eyes).

REVERSIBLE EFFECTS CLASS

An eye irritant Class 2 (irritating to eyes) is a test material that produces:

- at least in 2 of 3 tested animals a positive response of:

  corneal opacity ≥ 1 and/or
  iritis ≥ 1, and/or
  conjunctival redness≥ 2

  conjunctival edema (chemosis) ≥ 2

  calculated as the mean scores following grading at 24, 48 and 72 hours after installation of the test material, and

- which fully reverses within an observation period of normally 21 days

Within this category an eye irritant is considered **mildly irritating to eyes (Class 2A)** when the effects listed above are fully reversible within 7 days of observation.

29. For those chemicals where there is pronounced variability among animal responses, this information may be taken into account in determining the classification.

HARMONIZED SYSTEM FOR THE CLASSIFICATION OF CHEMICALS WHICH CAUSE RESPIRATORY OR SKIN SENSITISATION

PURPOSE, BASIS AND APPLICABILITY

1. The purpose of the harmonised criteria for classification of respiratory and dermal sensitisers is to give a common ground, which could be used internationally, for the hazard classification of sensitising properties of chemicals.

2. The basis for the harmonised criteria are those criteria which are currently in use in the OECD countries. Elements from these were integrated so as to maintain a high level of protection and to form harmonised criteria which could be agreed upon.

3. The criteria should be applicable on the hazard classification of chemicals irrespective of their end use.

I. RESPIRATORY SENSITISERS

Definitions

4. A respiratory sensitiser is a substance that will induce hypersensitivity of the airways following inhalation of the substance.

Classification Criteria

5. Substances shall be classified as respiratory sensitisers in accordance with the criteria given below:

- if there is evidence in humans that the substance can induce specific respiratory hypersensitivity**,** and/or
- where there are positive results from an appropriate animal test.

RATIONALE FOR THE SYSTEM

Human evidence

6. Evidence that a substance can induce specific respiratory hypersensitivity will normally be based on human experience. In this context, hypersensitivity is normally seen as asthma, but other hypersensitivity reactions such as rhinitis/conjunctivitis and alveolitis are also considered. The condition will have the clinical character of an allergic reaction. However, immunological mechanisms do not have to be demonstrated.

7. When considering the human evidence, it is necessary for a decision on classification to take into account in addition to the evidence from the cases:

- the size of the population exposed
- the extent of exposure.

8. The evidence referred to above could be

- clinical history and data from appropriate lung function tests related to exposure to the substance, confirmed by other supportive evidence which may include:
  - in vivo immunological test (e.g. skin prick test)
  - in vitro immunological test (e.g. serological analysis)
  - studies that may indicate other specific hypersensitivity reactions where immunological mechanisms of action have not been proven, e.g. repeated low-level irritation, pharmacologically mediated effects
  - a chemical structure related to substances known to cause respiratory hypersensitivity
- data from positive bronchial challenge tests with the substance conducted according to accepted guidelines for the determination of a specific hypersensitivity reaction.

9. Clinical history should include both medical and occupational history to determine a relationship between exposure to a specific substance and development of respiratory hypersensitivity. Relevant information includes aggravating factors both in the home and workplace, the onset and progress of the disease, family history and medical history of the patient in question. The medical history should also include a note of other allergic or airway disorders from childhood, and smoking history.

10. The results of positive bronchial challenge tests are considered to provide sufficient evidence for classification on their own. It is however recognised that in practice many of the examinations listed above will already have been carried out.

Animal studies

11. Data from appropriate animal studies which may be indicative of the potential of a substance to cause sensitisation by inhalation in humans may include:

- measurements of IgE and other specific immunological parameters, for example in mice
- specific pulmonary responses in guinea pigs.

EXPLANATORY NOTES

12. The mechanisms by which substances induce symptoms of asthma are not yet fully known. For preventative reasons these substances are considered as respiratory sensitisers. However, if on the basis of the evidence mentioned in paragraph 8, it can be demonstrated that these substances induce symptoms of

asthma by irritation only in people with bronchial hyperreactivity, they should not be considered as respiratory sensitisers.

13. At present recognised animal models for the testing of respiratory hypersensitivity are not available. Under certain circumstances, animal testing may be used, e.g. a modification of the guinea pig maximisation test for determination of relative allergenicity of proteins. However, these tests still need further validation.

14. Some substances causing respiratory sensitisation may in addition cause immunological contact urticaria and therefore should be considered for classification as a contact sensitisers (see part II).

II. CONTACT SENSITISERS

Definitions

15. A contact sensitiser is a substance that will induce an allergic response following skin contact.

Classification Criteria

16. Substances shall be classified as contact sensitisers in accordance with the criteria given below:

- if there is evidence in humans that the substance can induce sensitisation by skin contact in a substantial number of persons, or
- where there are positive results from an appropriate animal test.

RATIONALE FOR THE SYSTEM

17. For classification of a substance evidence should include any or all of the following:

- Positive data from patch testing, normally obtained in more than one dermatology clinic.
- Epidemiological studies showing allergic contact dermatitis caused by the substance. Situations in which a high proportion of those exposed exhibit characteristic symptoms are to be looked at with special concern, even if the number of cases is small.
- Positive data from appropriate animal studies.
- Positive data from experimental studies in man. (see General Considerations, paragraph 21).
- Well documented episodes of allergic contact dermatitis, normally obtained in more than one dermatology clinic.

18. Positive effects seen in either humans or animals will normally justify classification. Evidence from animal studies is usually much more reliable than evidence from human exposure. However, in cases where evidence is available from both sources, and there is conflict between the results, the quality and reliability of the evidence from both sources must be assessed in order to resolve the question of classification on a case-by-case basis. Normally, human data are not generated in controlled experiments with volunteers for the purpose of hazard classification but rather as part of risk assessment to confirm lack of effects seen in animal tests. Consequently, positive human data on contact sensitisation are usually derived from case-control or other, less defined studies. Evaluation of human data must therefore be carried out with caution as the frequency of cases reflect, in addition to the inherent properties of the substances, factors such as the exposure situation, bioavailability, individual predisposition and preventive measures taken. Negative human data should not normally be used to negate positive results from animal studies.

19. If none of the above mentioned conditions are met the substance need not be classified as a contact sensitiser. However, a combination of two or more indicators of contact sensitisation as listed below may alter the decision. This shall be considered on a case-by-case basis.

- Isolated episodes of allergic contact dermatitis.

- Epidemiological studies of limited power, e.g. where chance, bias or confounders have not been ruled out fully with reasonable confidence.

- Data from animal tests, performed according to existing guidelines, which do not meet the criteria given in the section on animal studies but are sufficiently close to the limit to be considered significant.

- Positive data from non-standard methods.

- Positive results from close structural analogues.

EXPLANATORY NOTES

Immunological Contact Urticaria

20. Substances meeting the criteria for classification as respiratory sensitisers may in addition cause immunological contact urticaria. Consideration should be given to classify these substances also as contact sensitisers. Substances which cause immunological contact urticaria without meeting the criteria for respiratory sensitisers should also be considered for classification as contact sensitisers.

21. There is no recognised animal model available to identify substances which cause immunological contact urticaria. Therefore, classification will normally be based on human evidence which will be similar to that for skin sensitisation.

Animal Studies

22. When an adjuvant type test method for skin sensitisation is used, a response of at least 30% of the animals is considered as positive. For a non-adjuvant test method a response of at least 15% of the animals is considered positive. Test methods for skin sensitisation are described in the OECD Guideline 406 (the Guinea Pig Maximisation test and the Buehler guinea pig test). Other methods may be used provided that they are well-validated and scientific justification is given.

23. The mouse ear swelling test, MEST, and the local lymph node assay, LLNA, appear to be reliable screening tests to detect moderate to strong sensitisers. The LLNA or the MEST can be used as a first stage in the assessment of skin sensitisation potential. In case of a positive result in either assay it may not be necessary to conduct a further guinea pig test.

24. When evaluating animal data, produced by testing according to the OECD or equivalent Guidelines for skin sensitisation, the rate of sensitised animals may be considered. This rate reflects the sensitising capacity of a substance in relation to its mildly irritating dose. This dose may vary between substances. A more appropriate evaluation of the sensitising capacity of a substance could be carried out if the dose-response relationship was known for the substance. This is an area that needs further development.

25. There are substances that are extremely sensitising at low doses where others require high doses and long time of exposure for sensitisation. For the purpose of hazard classification it may be preferable to distinguish between strong and moderate sensitisers. However, at present animal or other test systems to subcategorise sensitisers have not been validated and accepted. Therefore, subcategorisation should not yet be considered as part of the harmonised classification system. (See Background Information).

APPENDIX: BACKGROUND INFORMATION

1. Categorisation of sensitisers accounting for differences in sensitising capacity among substances would be a useful concept to develop. It may be appropriate to allocate both respiratory and dermal sensitizers to, for example, one of the following categories:

Category 1, Strong Sensitizer:

A strong sensitizer would be indicated by

- a high frequency of occurrence and/or severity of occurrence within an exposed population or
- a probability of occurrence of a high sensitization rate in humans based on animal or other tests.

Category 2, Sensitizer:

A low to moderate sensitizer would be indicated by

- a low or moderate frequency or severity of occurrence within an exposed population or
- a probability of occurrence of a low to moderate sensitization rate in humans based on animal or other tests.

2. Some authorities currently categorise strong sensitizers. However, at present, animal or other test systems to subcategorise sensitizers as indicated above, have not been validated and accepted. Work is going on to develop such models for the potency evaluation of contact allergens.

HARMONIZED SYSTEM FOR THE CLASSIFICATION OF CHEMICALS WHICH CAUSE MUTATIONS IN GERM CELLS

PURPOSE, BASIS AND APPLICABILITY

1. The purpose of the harmonized scheme for the classification of chemicals which may cause heritable mutations in germ cells in humans is to provide a common ground which could be used internationally for the classification of mutagens. All tests conducted according to validated and internationally accepted test guidelines are acceptable for the purpose of classifying substances.

2. To arrive at that classification scheme, test results are considered from experiments determining mutagenic and/or genotoxic effects in germ and/or somatic cells of exposed animals. Mutagenic and/or genotoxic effects determined in *in vitro* tests may also be considered.

3. The system is hazard based, classifying chemicals on the basis of their intrinsic ability to induce mutations in germ cells. The scheme is, therefore, not meant for the (quantitative) risk assessment of chemical substances.

DEFINITIONS

4. The classification system is primarily concerned with chemicals which may cause mutations in the germ cells of humans and these mutations can be transmitted to the progeny. However, mutagenicity/genotoxicity tests *in vitro* and in mammalian somatic cells *in vivo* will also be considered in the sub-divisions of the classification system.

5. In the present context, commonly found definitions of the terms mutagenic, mutagen, mutations and genotoxic are used, and a mutation is defined here as a permanent change in the amount or structure of the genetic material in a cell.

6. The term “mutation” applies both for heritable genetic changes that may be manifested at the phenotypic level, and for the underlying DNA modifications when known (including, for example, specific base pair changes and chromosomal translocations). The term “mutagenic” and “mutagen” will be used for agents giving rise to an increased occurrence of mutations in populations of cells and/or organisms.

7. The more general terms “genotoxic” and “genotoxicity” apply to agents or processes which alter the structure, information content, or segregation of DNA, including those which cause DNA damage by interfering with normal replication processes, or which in a non-physiological manner (temporarily) alter its replication. Genotoxicity test results are usually taken as indicators for mutagenic effects.

CLASSIFICATION CATEGORIES AND CRITERIA

8. The classification system comprises two different classes of germ cell mutagens to accommodate the weight of evidence available. The two-class system is described in the following.

Class 1:

Description: **Chemicals known to induce heritable mutations or to be regarded as if they induce heritable mutations in the germ cells of humans.**

Class 1a:

Description: Chemicals known to induce heritable mutations in germ cells of humans

Criteria: Positive evidence from human epidemiological studies.

Class 1b:

Description: Chemicals which should be regarded as if they induce heritable mutations in the germ cells of humans.

Criteria:

- Positive result(s) from *in vivo* heritable germ cell mutagenicity tests in mammals; or
- Positive result(s) from *in vivo* somatic cell mutagenicity tests in mammals, in combination with some evidence that the substance has potential to cause mutations to germ cells. This supporting evidence may, for example, be derived from mutagenicity/genotoxic tests in germ cells *in vivo*, or by demonstrating the ability of the substance or its metabolite(s) to interact with the genetic material of germ cells; or
- Positive results from tests showing mutagenic effects in the germ cells of humans, without demonstration of transmission to progeny; for example, an increase in the frequency of aneuploidy in sperm cells of exposed people.

Class 2:

Description: **Chemicals which cause concern for man owing to the possibility that they may induce heritable mutations in the germ cells of humans.**

Criteria: Positive evidence obtained from experiments in mammals and/or in some cases from *in vitro* experiments, obtained from:

- Somatic cell mutagenicity tests *in vivo*, in mammals; or
- Other *in vivo* somatic cell genotoxicity tests which are to be supported by positive results from *in vitro* mutagenicity assays.

Nota Bene:

- Chemicals which are positive in *in vitro* mammalian mutagenicity assays, and which also show chemical structure activity relationship to known germ cell mutagens, should be considered for classification as class 2 mutagens.

RATIONALE FOR THE PROPOSED SYSTEM

9. Classification for heritable effects in human germ cells is made on the basis of well conducted, sufficiently validated tests, preferably as described in OECD Test Guidelines. Evaluation of the test results should be done using expert judgement and all the available evidence should be weighed for classification.

10. Examples of *in vivo* heritable germ cell mutagenicity tests are:

Rodent dominant lethal mutation test (OECD 478)
Mouse heritable translocation assay (OECD 485)
Mouse specific locus test

11. Examples of in vivo somatic cell mutagenicity tests are:

Mammalian bone marrow micronucleus test (OECD 474)
Mammalian bone marrow chromosome aberration test (OECD 475)
Mouse spot test (OECD 484)
Mammalian erythrocyte micronucleus test (OECD 474)

12. Examples of mutagenicity/genotoxicity tests in germ cells are:

A) Mutagenicity tests:
Mammalian spermatogonial chromosome aberration test (OECD 483)
Spermatid micronucleus assay

B) Genotoxicity tests:
Sister chromatid exchange analysis in spermatogonia
Unscheduled DNA synthesis test (UDS) in testicular cells

13. Examples of genotoxicity tests in somatic cells are:

Liver Unscheduled DNA Synthesis (UDS) in vivo (OECD 486)
Mammalian bone marrow sister chromatid exchanges (SCE)

14. Examples of in vitro mutagenicity tests are:

In vitro mammalian chromosome aberration test (OECD 473)
In vitro mammalian cell gene mutation test (OECD 476)
Bacterial reverse mutation tests (OECD 471)

15. The classification of individual substances should be based on the total weight of evidence available, using expert judgement. In those instances where a single well-conducted test is used for classification, it should provide clear and unambiguously positive results. If new, well validated, tests arise these may also be used in the total weight of evidence to be considered. The relevance of the route of exposure used in the study of the chemical compared to the route of human exposure should also be taken into account.

EXPLANATORY NOTES

16. It becomes increasingly clear that the process of chemical-induced tumorigenesis in man and animals involves (an accumulation of) genetic changes in proto-oncogenes and/or tumour suppressor genes of somatic cells. Therefore, the demonstration of mutagenic properties of chemicals in somatic and/or germ cells of mammals *in vivo* may have implications for the potential classification of these chemicals as carcinogens (cf. chapter "Harmonization of Classification Systems on Carcinogens").

HARMONIZED SYSTEM FOR THE CLASSIFICATION OF CHEMICALS WHICH CAUSE CANCER

PURPOSE, BASIS AND APPLICABILITY

1. The purpose of the harmonized system for the classification of chemicals which may cause cancer is to provide common ground which could be used internationally for the classification of carcinogenic substances.

2. The scheme is applicable to the classification of all chemicals. The system deals only with chemical substances. The application to classification of preparations/products/mixtures should be considered as a further step by the Working Group on Mixtures.

DEFINITIONS

3. The term "carcinogen" denotes a chemical substance or a mixture of chemical substances which induce cancer or increase its incidence. Substances which have induced benign and malignant tumours in well performed experimental studies on animals are considered also to be presumed or suspected human carcinogens unless there is strong evidence that the mechanism of tumour formation is not relevant for humans.

4. Classification of a chemical as posing *a* carcinogenic hazard is based on the inherent properties of the substance and does not provide information on the level of the human cancer risk which the use of the chemical may represent.

CLASSIFICATION CATEGORIES AND CRITERIA

5. For the purpose of classification for carcinogenicity, chemical substances are allocated to one of two classes based on strength of evidence and additional considerations (weight of evidence). In certain instances route specific classification may be warranted.

CLASS 1: KNOWN OR PRESUMED HUMAN CARCINOGENS

The placing of a chemical in Class 1 is done on the basis of epidemiological and/or animal data. An individual chemical may be further distinguished:

Class 1A: KNOWN to have carcinogenic potential for humans; the placing of a chemical is largely based on human evidence.

Class 1B: **PRESUMED** to have carcinogenic potential for humans; the placing of a chemical is largely based on animal evidence.

Based on strength of evidence together with additional considerations**,** such evidence may be derived from human studies that establish a causal relationship between human exposure to a chemical and the development of cancer (known human carcinogen). Alternatively, evidence may be derived from animal experiments for which there is sufficient evidence to demonstrate animal carcinogenicity (presumed human carcinogen). In addition**,** on a case by case basis, scientific judgement may warrant a decision of presumed human carcinogenicity derived from studies showing limited evidence of carcinogenicity in humans together with limited evidence of carcinogenicity in experimental animals.

Classification: Class 1 (A and B) Carcinogen

CLASS 2: SUSPECTED HUMAN CARCINOGENS

The placing of a chemical in Class 2 is done on the basis of evidence obtained from human and/or animal studies, but which is not sufficiently convincing to place the chemical in Class 1.

Based on strength of evidence together with additional considerations, such evidence may be from either limited evidence of carcinogenicity in human studies or from limited evidence of carcinogenicity in animal studies.

Classification: Class 2 Carcinogen

RATIONALE FOR THE PROPOSED SYSTEM

6. **Classification as Carcinogen** is made on the basis of evidence from reliable and acceptable methods, and is intended to be used for chemicals which have an intrinsic property to produce such toxic effects. The evaluations should be based on all existing data, peer-reviewed published studies and additional data accepted by regulatory agencies.

7. **Carcinogen classification** is a one-step, criterion-based process that involves two interrelated determinations: evaluations of strength of evidence and consideration of all other relevant information to place chemicals with human cancer potential into hazard classes.

8. **Strength of evidence** involves the enumeration of tumours in human and animal studies and determination of their level of statistical significance. Sufficient human evidence demonstrates causality between human exposure and the development of cancer, whereas sufficient evidence in animals shows a causal relationship between the agent and an increased incidence of tumours. Limited evidence in humans is demonstrated by a positive association between exposure and cancer, but a causal relationship cannot be stated. Limited evidence in animals is provided when data suggest a carcinogenic effect, but are less than sufficient. The terms "sufficient", and "limited" are used here as they have been defined by the International Agency for Research on Cancer (IARC) and are cited in the Background Information for this document.

9. **Additional considerations** (weight of evidence). Beyond the determination of the strength of evidence for carcinogenicity, a number of other factors should be considered that influence the overall likelihood that an agent may pose a carcinogenic hazard in humans. The full list of factors that influence this determination is very lengthy, but some of the important ones are considered here.

10. The factors can be viewed as either increasing or decreasing the level of concern for human carcinogenicity. The relative emphasis accorded to each factor depends upon the amount and coherence of evidence bearing on each. Generally there is a requirement for more complete information to decrease than to increase the level of concern. Additional considerations should be used in evaluating the tumour findings and the other factors in a case-by-case manner.

11. Some important factors which may be taken into consideration, when assessing the overall level of concern are:

- Tumor type and background incidence.
- Multisite responses.
- Progression of lesions to malignancy.
- Reduced tumor latency.

12. Additional factors on which the evaluation may increase or decrease the level of concern include:

- Whether responses are in single or both sexes.
- Whether responses are in a single species or several species.
- Structural similarity or not to a chemical(s) for which there is good evidence of carcinogenicity.
- Routes of exposure.
- Comparison of absorption, distribution, metabolism and excretion between test animals and humans.
- The possibility of a confounding effect of excessive toxicity at test doses.
- Mode of action and its relevance for humans, such as mutagenicity, cytotoxicity with growth stimulation, mitogenesis, immunosuppression.

13. **Mutagenicity.** It is recognised that genetic events are central in the overall process of cancer development. Therefore evidence of mutagenic activity *in vivo* may indicate that a chemical has a potential for carcinogenic effects.

EXPLANATORY NOTES

14. The following additional considerations apply to classification of chemicals into either Class 1 or Class 2. A chemical that has not been tested for carcinogenicity may in certain instances be classified in

Class 1 or Class 2 based on tumour data from a structural analogue together with substantial support from consideration of other important factors such as formation of common significant metabolites, e.g. for benzidine congener dyes.

15. The classification should take into consideration whether or not the chemical is absorbed by a given route(s); or whether there are only local tumours at the site of administration for the tested route(s), and adequate testing by other major route(s) show lack of carcinogenicity.

16. It is important that whatever is known of the physico-chemical, toxicokinetic and toxicodynamic properties of the substances, as well as any available relevant information on chemical analogues, i.e. structure activity relationship, is taken into consideration when undertaking classification.

17. It is realised that some regulatory authorities may need flexibility beyond that developed in the hazard classification scheme. For inclusion into Safety Data Sheets positive results in any carcinogenicity study performed according to good scientific principles with statistically significant results may be considered.

18. Guidance on the importance of the different factors mentioned in paragraph 12 has to be elaborated in order to indicate their effects or level of concern.

19. The relative hazard potential of a chemical is a function of its intrinsic potency. There is great variability in potency among chemicals, and it may be important to account for these potency differences. The work that remains to be done is to examine methods for potency estimation. Carcinogenic potency as used here does not preclude risk assessment. (See Background Information)

20. The proceedings of the recent WHO/IPCS working group to harmonized risk assessment for carcinogenicity points to a number of scientific questions arising for classification of chemicals e.g. mouse liver tumours, peroxisome proliferation, receptor-mediated reactions, chemicals which are carcinogenic only at toxic doses and which do not demonstrate mutagenicity. Accordingly, there is a need to articulate the principles necessary to resolve these scientific issues which have led to diverging classifications in the past. Once these issues are resolved, there would be a firm foundation for classification of a number of chemical carcinogens.

21. Data already generated for classifying chemicals under existing systems should be acceptable when reviewing these chemicals with regard to classification under the harmonized system. Further testing should not (normally) be necessary.

APPENDIX : BACKGROUND INFORMATION

I. Evaluation of the Strength of Evidence for Carcinogenicity Arising from Human and Experimental Data Adopted by the International Agency for Research on Cancer (IARC)

Carcinogenicity in humans

1. The evidence relevant to carcinogenicity from studies in humans is classified into one of the following categories:

- **Sufficient evidence of carcinogenicity:** The Working Group considers that a causal relationship has been established between exposure to the agent, mixture or exposure circumstance and human cancer. That is, a positive relationship has been observed between

exposure and cancer in studies in which chance, bias and confounding could be ruled out with reasonable confidence.

- **Limited evidence of carcinogenicity:** A positive association has been observed between exposure to the agent, mixture or exposure circumstance and cancer for which a causal interpretation is considered by the Working Group to be credible, but chance, bias or confounding could not be ruled out with reasonable confidence.

2. In some instances the above categories may be used to classify the degree of evidence related to carcinogenicity in specific organs or tissues.

Carcinogenicity in experimental animals

3. The evidence relevant to carcinogenicity in experimental animals is classified into one of the following categories:

- **Sufficient evidence of carcinogenicity:** The Working Group considers that a causal relationship has been established between the agent and an increased incidence of malignant neoplasms or of an appropriate combination of benign and malignant neoplasms in (a) two or more species of animals or (b) in two or more independent studies in one species carried out at different times or in different laboratories or under different protocols.

- Exceptionally, a single study in one species might be considered to provide sufficient evidence of carcinogenicity when malignant neoplasms occur to an unusual degree with regard to incidence, site, type of tumour or age at onset.

- **Limited evidence of carcinogenicity:** The data suggest a carcinogenic effect but are limited for making a definitive evaluation because, e.g., (a) the evidence of carcinogenicity is restricted to a single experiment; or (b) there are unresolved questions regarding the adequacy of the design, conduct or interpretation of the study; or (c) the agent or mixture increases the incidence only of benign neoplasms or lesions of uncertain neoplastic potential, or of certain neoplasms which may occur spontaneously in high incidences in certain strains.

II. Considerations of Potency for Labelling Limits

4. The considerations as laid out below were excerpted from the Report of the Meeting of the Working Group on Harmonization of Classification and Labelling of Carcinogens, Washington, DC, 17-18 October 1995.

Purpose

5. The purpose of establishing a potency scheme to be used for labelling of substances, preparations (mixtures) and contaminants is to provide for practical minimum levels of carcinogens in substances for which labelling would be required. It will result in labelling highly potent materials more strictly and less potent materials less strictly. A further purpose is to eliminate unnecessary labelling. In addition, use of a potency scheme may encourage risk reduction through purification of chemical substances or reformulating preparations.

Background

6. A large number of chemicals have been classified as carcinogenic and placed into various categories for labelling or other regulatory purpose. Chemicals that have been identified as carcinogenic may also occur as components of preparations (mixtures), impurities or additives. Gold and co-authors (Environ Health Perspect 79: 259, 1989) calculated doses from animal testing which result in tumours in half the dosed animals (TD50 values span a range of more than eight orders of magnitude). Most classification systems do not take into account the wide range of potencies of these chemicals.

7. Carcinogens are in some countries divided into three potency groups: high, medium and low. Potency is in these instances determined using dose-response data in the observed dosing range for laboratory animals. Additional indicators of potency such as tumour site and species specificity, or species differences in toxicokinetics may also be used. Such potency groups are used to set upper limits for the classification of substances as carcinogens and for the purpose of initiating labelling. They have also been used for the classification and determination of labelling provisions for preparations (mixtures) of carcinogenic chemicals.

8. Some countries have implemented a scheme where 0.1% is used as a default limit value for labelling of substances and preparations (mixtures) as carcinogens with sufficient data for carcinogenicity. In these countries chemicals with medium carcinogenic potency are labelled if they occur in chemical substances at or above this level. Many carcinogenic compounds fall into the medium range. Carcinogens with high potency might be classified and labelled at lower levels and carcinogens with low potency could be classified and labelled only when they occur at higher levels. Some countries use 1% as a default limit value for low potency carcinogens and for carcinogens with more limited data.

9. Some regulatory authorities do not have the obligation to perform potency determinations. If a chemical carcinogen is a candidate for a potency rating outside of the default range, such chemicals should be referred to an international group for its determination.

Observations

10. The Working Group agreed that it would be useful to explore further the concept of using potency to make labelling decisions. Initial thoughts of the Working Group are presented here.

11. Potency ranking of carcinogens should not be determined or refined more precisely than by ten-fold factors in light of differences in species response, tumour types and the limits of standardization of test protocols. In light of these points, a scheme for classification and labelling purposes which separates carcinogens into potency groupings serves the practical purposes listed above.

12. The use of potency for establishing limits does not preclude the ability of authorities to perform quantitative risk assessments of exposures to carcinogenic substances for regulatory purposes.

13. Potency determinations should be based on well performed studies which are peer reviewed, performed according to good laboratory practices, or are deemed acceptable by regulatory authorities.

HARMONIZED SYSTEM FOR THE CLASSIFICATION OF CHEMICALS WHICH CAUSE REPRODUCTIVE TOXICITY

PURPOSE, BASIS, AND APPLICABILITY

1. The purpose of the harmonized system for the classification of chemicals which may cause an adverse effect on reproduction in humans is to provide a common ground which could be used internationally for the classification of reproductive toxicants.

2. The system is hazard based, classifying chemicals on the basis of intrinsic ability to produce an adverse effect on reproductive function or capacity, and/or on development of the offspring.

3. The present system involves consideration of any substance-related adverse effect on reproduction seen in humans, or observed in appropriate tests conducted in experimental animals.

4. The Explanatory Notes provide essential guidance and should be regarded as an integral part of the Classification System.

REPRODUCTIVE TOXICITY: DEFINITIONS

5. Reproductive toxicity includes adverse effects on sexual function and fertility in adult males and females, as well as developmental toxicity in the offspring. The definitions presented below are adapted from those agreed at the IPCS/OECD Workshop for the Harmonisation of Risk Assessment for Reproductive and Developmental Toxicity, Carshalton, UK, 17-21 October, 1994 (OECD Monograph Series on Testing and Assessment No. 17, 1998). For classification purposes, the known induction of genetically-based inheritable effects in the offspring is addressed elsewhere, since in the present classification system it is considered more appropriate to address such effects under the separate end-point of germ-cell mutagenicity.

6. In this classification system, reproductive toxicity is subdivided under two main headings:

a) Adverse effects on reproductive ability or capacity

7. Any effect of chemicals that would interfere with reproductive ability or capacity. This may include, but not be limited to, alterations to the female and male reproductive system, adverse effects on onset of puberty, gamete production and transport, reproductive cycle normality, sexual behaviour, fertility, parturition, premature reproductive senescence, or modifications in other functions that are dependent on the integrity of the reproductive systems.

8. Adverse effects on or via lactation can also be included in reproductive toxicity, but for classification purposes, such effects are treated separately (see paragraph 16). This is because it is desirable to be able to classify chemicals specifically for adverse effect on lactation so that a specific hazard warning about this effect can be provided for lactating mothers.

b) Adverse effects on development of the offspring

9. Taken in its widest sense, developmental toxicity includes any effect which interferes with normal development of the conceptus, either before or after birth, and resulting from exposure of either parent prior to conception, or exposure of the developing offspring during prenatal development, or postnatally, to the time of sexual maturation.

10. However, it is considered that classification under the heading of developmental toxicity is primarily intended to provide hazard warning for pregnant women and men and women of reproductive capacity. Therefore, for pragmatic purposes of classification, developmental toxicity essentially means adverse effects induced during pregnancy, or as a result of parental exposure. These effects can be manifested at any point in the life span of the organism. The major manifestations of developmental toxicity include (1) death of the developing organism, (2) structural abnormality, (3) altered growth, and (4) functional deficiency.

CLASSIFICATION

Weight of Evidence

11. Classification as a reproductive toxicant is made on the basis of an assessment of the total weight of evidence. This means that all available information that bears on the determination of reproductive toxicity is considered together. Included are such information as epidemiological studies and case reports in humans and specific reproduction studies along with sub-chronic, chronic and special study results in animals that provide relevant information regarding toxicity to reproductive and related endocrine organs. Evaluation of substances chemically related to the material under study may also be included, particularly when information on the material is scarce. The weight given to the available evidence will be influenced by factors such as the quality of the studies, consistency of results, nature and severity of effects, level of statistical significance for intergroup differences, number of endpoints affected, relevance of route of administration to humans and freedom from bias. Both positive and negative results are assembled together into a weight of evidence determination. However, a single, positive study performed according to good scientific principles and with statistically or biologically significant positive results may justify classification (see also paragraph 13).

12. Toxicokinetic studies in animals and humans, site of action and mechanism or mode of action study results may provide relevant information, which could reduce or increase concerns about the hazard to human health. If it can be conclusively demonstrated that the clearly identified mechanism or mode of action has no relevance for humans or when the toxicokinetic differences are so marked that it is certain that the hazardous property will not be expressed in humans then a substance which produces an adverse effect on reproduction in experimental animals should not be classified.

13. In some reproductive toxicity studies in experimental animals the only effects recorded may be considered of low or minimal toxicological significance and classification may not necessarily be the outcome. These include for example small changes in semen parameters or in the incidence of spontaneous defects in the foetus, small changes in the proportions of common fetal variants such as are observed in skeletal examinations, or in fetal weights, or small differences in postnatal developmental assessments.

14. Data from animal studies ideally should provide clear evidence of specific reproductive toxicity in the absence of other, systemic, toxic effects. However, if developmental toxicity occurs together with other toxic effects in the dam, the potential influence of the generalised adverse effects should be assessed to the extent possible. The preferred approach is to consider adverse effects in the embryo/fetus first, and then evaluate maternal toxicity, along with any other factors which are likely to have influenced these effects, as part of the weight of evidence. In general, developmental effects that are observed at maternal toxic doses should not be automatically discounted. Discounting developmental effects that are observed at maternal toxic doses can only be done on a case-by-case basis when a causal relationship is established or refuted.

15. If appropriate information is available it is important to try to determine whether developmental toxicity is due to a specific maternally mediated mechanism or to a non-specific secondary mechanism, like maternal stress and the disruption of homeostasis. Generally, the presence of maternal toxicity should not be used to negate findings of embryo/fetal effects , unless it can be clearly demonstrated that the effects are secondary non-specific effects. This is especially the case when the effects in the offspring are significant, e.g. irreversible effects such as structural malformations. In some situations it is reasonable to assume that reproductive toxicity is due to a secondary consequence of maternal toxicity and discount the effects, for example if the chemical is so toxic that dams fail to thrive and there is severe inanition; they are incapable of nursing pups; or they are prostrate or dying.

Hazard classes

16. For the purpose of classification for reproductive toxicity, chemical substances are allocated to one of two classes. Effects on reproductive ability or capacity, and on development, are considered as separate issues.

Class 1: KNOWN OR PRESUMED HUMAN REPRODUCTIVE OR DEVELOPMENTAL TOXICANT

This Class includes substances which are known to have produced an adverse effect on reproductive ability or capacity or on development in humans or for which there is evidence from animal studies, possibly supplemented with other information, to provide a strong presumption that the substance has the capacity to interfere with reproduction in humans. For regulatory purposes, a substance can be further distinguished on the basis of whether the evidence for classification is primarily from human data (Class 1A) or from animal data (Class 1B).

Class 1A: KNOWN to have produced an adverse effect on reproductive ability or capacity or on development in humans. The placing of the substance in this class is largely based on evidence from humans.

Class 1B: PRESUMED to produce an adverse effect on reproductive ability or capacity or on development in humans. The placing of the substance in this class is largely based on evidence from experimental animals. Data from animal studies should provide clear evidence of specific reproductive toxicity in the absence of other toxic effects, or if occurring together with other toxic effects the adverse effect on reproduction is considered not to be a secondary non-specific consequence of other toxic effects. However, when there is mechanistic information that raises doubt about the relevance of the effect for humans, classification in Class 2 may be more appropriate.

Class 2: SUSPECTED HUMAN REPRODUCTIVE OR DEVELOPMENTAL TOXICANT

This Class includes substances for which there is some evidence from humans or experimental animals, - possibly supplemented with other information - of an adverse effect on reproductive ability or capacity, or on development, in the absence of other toxic effects, or if occurring together with other toxic effects the adverse effect on reproduction is considered not to be a secondary non-specific consequence of the other toxic effects, and where the evidence is not sufficiently convincing to place the substance in Class 1. For instance, deficiencies in the study may make the quality of evidence less convincing, and in view of this Class 2 could be the more appropriate classification.

EFFECTS ON OR VIA LACTATION

Effects on or via lactation are allocated to a separate single class. It is appreciated that for many substances there is no information on the potential to cause adverse effects on the offspring via lactation. However, for substances which are absorbed by women and have been shown to interfere with lactation or which may be present (including metabolites) in breast milk in amounts sufficient to cause concern for the health of a breastfed child, should be classified to indicate this property hazardous to breastfed babies. This classification can be assigned on the basis of:

(a) absorption, metabolism, distribution and excretion studies that would indicate the likelihood the substance would be present in potentially toxic levels in breast milk; and/or

(b) results of one or two generation studies in animals which provide clear evidence of adverse effect in the offspring due to transfer in the milk or adverse effect on the quality of the milk; and/or

(c) human evidence indicating a hazard to babies during the lactation period.

BASIS OF CLASSIFICATION

17. Classification is made on the basis of the appropriate criteria, outlined above, and an assessment of the total weight of evidence. Classification as a reproductive or developmental toxicant is intended to be used for chemicals which have an intrinsic, specific property to produce an adverse effect on reproduction or development and chemicals should not be so classified if such an effect is produced solely as a non-specific secondary consequence of other toxic effects.

18. In the evaluation of toxic effects on the developing offspring, it is important to consider the possible influence of maternal toxicity.

19. For human evidence to provide the primary basis for a Class 1A classification there must be reliable evidence of adverse effect on reproduction in humans. Evidence used for classification should ideally be from well conducted epidemiological studies which include the use of appropriate controls, balanced assessment, and due consideration of bias or confounding factors. Less rigorous data from studies in humans should be supplemented with adequate data from studies in experimental animals and classification in Class 1B should be considered.

20. Data already generated for classifying chemicals under existing systems should be acceptable when reviewing these chemicals with regard to classification under the harmonised system. Further testing should not normally be necessary.

EXPLANATORY NOTES

Maternal toxicity

21. Development of the offspring throughout gestation and during the early post-natal stages can be influenced by toxic effects in the mother either through non-specific mechanisms related to stress and the disruption of maternal homeostasis, or by specific maternally-mediated mechanisms. So, in the interpretation of the developmental outcome to decide classification for developmental effects it is important to consider the possible influence of maternal toxicity. This is a complex issue because of uncertainties surrounding the relationship between maternal toxicity and developmental outcome. Expert judgement and a weight of evidence approach, using all available studies, should be used to determine the degree of influence that should be attributed to maternal toxicity when interpreting the criteria for classification for developmental effects. The adverse effects in the embryo/fetus should be first considered, and then maternal toxicity, along with any other factors which are likely to have influenced these effects, as weight of evidence, to help reach a conclusion about classification.

22. Based on pragmatic observation, it is believed, that maternal toxicity may, depending on severity, influence development via non-specific secondary mechanisms, producing effects such as depressed foetal weight, retarded ossification, and possibly resorptions and certain malformations in some strains of certain species. However, the limited number of studies which have investigated the relationship between developmental effects and general maternal toxicity have failed to demonstrate a consistent, reproducible relationship across species. Developmental effects which occur even in the presence of maternal toxicity are considered to be evidence of developmental toxicity, unless it can be unequivocally demonstrated on a case by case basis that the developmental effects are secondary to maternal toxicity. Moreover, classification should be considered where there is significant toxic effect in the offspring, e.g. irreversible effects such as structural malformations, embryo/fetal lethality, significant post-natal functional deficiencies.

23. Classification should not automatically be discounted for chemicals that produce developmental toxicity only in association with maternal toxicity, even if a specific maternally-mediated mechanism has been demonstrated. In such a case, classification in Class 2 may be considered more appropriate than Class 1. However, when a chemical is so toxic that maternal death or severe inanition results, or the dams are prostrate and incapable of nursing the pups, it may be reasonable to assume that developmental toxicity is produced solely as a secondary consequence of maternal toxicity and discount the developmental effects. Classification may not necessarily be the outcome in the case of minor developmental changes e.g. small reduction in fetal/pup body weight, retardation of ossification when seen in association with maternal toxicity.

24. Some of the end points used to assess maternal toxicity are provided below. Data on these end points, if available, needs to be evaluated in light of their statistical or biological significance and dose response relationship.

> Maternal Mortality: An increased incidence of mortality among the treated dams over the controls should be considered evidence of maternal toxicity if the increase occurs in a dose-related manner and can be attributed to the systemic toxicity of the test material. Maternal mortality greater than 10% is considered excessive and the data for that dose level should not normally be considered for further evaluation.

Mating Index (no. animals with seminal plugs or sperm/no. mated x 100)[1]
Fertility Index (no. animals with implants/no. of matings x 100)[1]
Gestation Length (if allowed to deliver)
Body Weight and Body Weight Change: Consideration of the maternal body weight change and/or adjusted (corrected) maternal body weight should be included in the evaluation of maternal toxicity whenever such data are available. The calculation of a adjusted (corrected) mean maternal body weight change, which is the difference between the initial and terminal body weight minus the gravid uterine weight (or alternatively, the sum of the weights of the foetuses), may indicate whether the effect is maternal or intrauterine. In rabbits, the body weight gain may not be useful indicators of maternal toxicity because of normal fluctuations in body weight during pregnancy.
Food and Water Consumption (if relevant): The observation of a significant decrease in the average food or water consumption in treated dams compared to the control group may be useful in evaluating maternal toxicity, particularly when the test material is administered in the diet or drinking water. Changes in food or water consumption should be evaluated in conjunction with maternal body weights when determining if the effects noted are reflective of maternal toxicity or more simply, unpallatibility of the test material in feed or water.
Clinical evaluations (including clinical signs, markers, hematology and clinical chemistry studies): The observation of increased incidence of significant clinical signs of toxicity in treated dams relative to the control group may be useful in evaluating maternal toxicity. If this is to be used as the basis for the assessment of maternal toxicity, the types, incidence, degree and duration of clinical signs should be reported in the study. Examples of frank clinical signs of maternal intoxication include: coma, prostration, hyperactivity, loss of righting reflex, ataxia, or laboured breathing.
Postmortem data: Increased incidence and/or severity of postmortem findings may be indicative of maternal toxicity. This can include gross or microscopic pathological findings or organ weight data, e.g., absolute organ weight, organ-to-body weight ratio, or organ-to-brain weight ratio. When supported by findings of adverse histopathological effects in the affected organ(s), the observation of a significant change in the average weight of suspected target organ(s) of treated dams, compared to those in the control group, may be considered evidence of maternal toxicity.

Potency and cut-off doses

25. In the present scheme, the relative potency of a chemical to produce a toxic effect on reproduction is not included in the criteria for reaching a conclusion regarding classification. Nevertheless, during the development of this scheme it was suggested that cut-off dose levels should be included, in order to provide some means of assessing and categorising the potency of chemicals for the ability to produce an adverse effect on reproduction. This concept has not been readily accepted by all member countries because of concerns that any specified cut-off level may be exceeded by human exposure levels in certain situations, e.g. inhalation of volatile solvents, the level may be inadequate in cases where humans are more sensitive than the animal model, and because of disagreements about whether or not potency is a component of hazard.

26. There has been interest in this concept to further consider it as a future development of the classification scheme.

1 . It is recognised that this index can also be affected by the male

Limit dose

27. Member countries appear to be in agreement about the concept of a limit dose, above which the production of an adverse effect may be considered to be outside the criteria which lead to classification. However, there is disagreement between members regarding the inclusion within the criteria of a specified dose as a limit dose. Some Test Guidelines specify a limit dose, other Test Guidelines qualify the limit dose with a statement that higher doses may be necessary if anticipated human exposure is sufficiently high that an adequate margin of exposure would not be achieved. Also, due to species differences in toxicokinetics, establishing a specific limit dose may not be adequate for situations where humans are more sensitive than the animal model.

28. In principle, adverse effects on reproduction seen only at very high dose levels in animal studies (for example doses that induce prostration, severe inappetence, excessive mortality) would not normally lead to classification, unless other information is available, e.g. toxicokinetics information indicating that humans may be more susceptible than animals, to suggest that classification is appropriate. Please also refer to the section on Maternal Toxicity for further guidance in this area.

29. However, specification of the actual 'limit dose' will depend upon the test method that has been employed to provide the test results, e.g. in the OECD Test Guideline for repeated dose toxicity studies by the oral route, an upper dose of 1000 mg/kg unless expected human response indicates the need for a higher dose level, has been recommended as a limit dose.

Animal and experimental data

30. A number of internationally accepted test methods are available; these include methods for developmental toxicity testing (e.g., OECD Test Guideline 414, ICH Guideline S5A, 1993), methods for peri- and post-natal toxicity testing (e.g. ICH S5B, 1995) and methods for one or two-generation toxicity testing (e.g. OECD Test Guidelines 415, 416).

31. Results obtained from Screening Tests (e.g. OECD Guidelines 421 - Reproduction/Developmental Toxicity Screening Test, and 422 - Combined Repeated Dose Toxicity Study with Reproduction/Development Toxicity Screening Test) can also be used to justify classification, although it is recognised that the quality of this evidence is less reliable than that obtained full studies.

32. Adverse effects or changes, seen in short- or long-term repeated dose toxicity studies, which are judged likely to impair reproductive ability or capacity and which occur in the absence of significant generalised toxicity, may be used as a basis for classification, e.g. histopathological changes in the gonads.

33. Evidence from in vitro assays, or non-mammalian tests, and from analogous substances using structure-activity relationship (SAR), can contribute to the procedure for classification. In all cases of this nature, expert judgement must be used to assess the adequacy of the data. Inadequate data should not be used as a primary support for classification.

34. It is preferable that animal studies are conducted using appropriate routes of administration which relate to the potential route of human exposure. However, in practice, reproductive toxicity studies are commonly conducted using the oral route, and such studies will normally be suitable for evaluating the hazardous properties of the substance with respect to reproductive toxicity. However, if it can be conclusively demonstrated that the clearly identified mechanism or mode of action has no relevance for

humans or when the toxicokinetic differences are so marked that it is certain that the hazardous property will not be expressed in humans then a substance which produces an adverse effect on reproduction in experimental animals should not be classified.

35. Studies involving routes of administration such as intravenous or intraperitoneal injection, which may result in exposure of the reproductive organs to unrealistically high levels of the test substance, or elicit local damage to the reproductive organs, e.g. by irritation, must be interpreted with extreme caution and on their own would not normally be the basis for classification.

HARMONIZED SYSTEM FOR THE CLASSIFICATION OF CHEMICALS WHICH CAUSE TARGET ORGAN ORIENTED SYSTEMIC TOXICITY

(to be inserted here after approval by the OECD Advisory Group on Harmonization of Classification and Labelling and subsequent endorsement by the Joint Meeting)

Purpose, basis and applicability

Definitions

Classification categories and criteria

Rationale for the proposed System

Explanatory notes

Literature

Background information

HARMONIZED SYSTEM FOR THE CLASSIFICATION OF CHEMICALS WHICH ARE HAZARDOUS FOR THE AQUATIC ENVIRONMENT

PURPOSE, BASIS AND APPLICABILITY

1. The harmonized system for classifying chemical substances for the hazards they present to the aquatic environment is based on a consideration of the existing systems listed below. The aquatic environment may be considered in terms of the aquatic organisms that live in the water, and the aquatic ecosystem of which they are part. To that extent, the proposal does not address aquatic pollutants for which there may be a need to consider effects beyond the aquatic environment such as the impacts on human health etc. The basis, therefore, of the identification of hazard is the aquatic toxicity of the substance, although this may be modified by further information on the degradation and bioaccumulation behaviour.

2. The proposed system is intended specifically for use with chemical substances and is not intended at this stage to cover preparations or other mixtures such as formulated pesticides. Its application to mixtures is deferred to the OECD Working Group on Mixtures. While the scheme is intended to apply to all substances, it is recognised that for some substances, e.g. metals, poorly soluble substances etc., special guidance will be necessary. A Guidance Document will thus be prepared to cover issues such as data interpretation and the application of the criteria defined below to such groups of substances. Considering the complexity of this endpoint and the breadth of the application of the system, the Guidance Document is considered an important element in the operation of the harmonised scheme.

3. Consideration has been given to existing classification systems as currently in use, including the EU Supply and Use Scheme, the revised GESAMP hazard evaluation procedure, IMO Scheme for Marine Pollutant, the European Road and Rail Transport Scheme (RID/ADR), the Canadian and US Pesticide systems and the US Land Transport Scheme. The harmonized scheme is considered suitable for use for packaged goods in both supply and use and multimodal transport schemes, and elements of it may be used for bulk land transport and bulk marine transport under MARPOL 73/78 Annex II insofar as this uses aquatic toxicity.

DEFINITIONS AND DATA REQUIREMENTS

4. The basic elements for use within the harmonized system are:
   - acute aquatic toxicity;
   - potential for or actual bioaccumulation;
   - degradation (biotic or abiotic) for organic chemicals; and
   - chronic aquatic toxicity.

5. While data from internationally harmonized test methods are preferred, in practice, data from national methods may also be used where they are considered as equivalent. In general, it has been agreed that freshwater and marine species toxicity data can be considered as equivalent data and are preferably to be derived using OECD Test Guidelines or equivalent according to the principles of GLP. Where such data are not available classification should be based on the best available data.

Acute toxicity

6. Acute aquatic toxicity would normally be determined using a fish 96 hour LC_{50} (OECD Test Guideline 203 or equivalent), a crustacea species 48 hour EC_{50} (OECD Test Guideline 202 or equivalent) and/or an algal species 72 or 96 hour EC_{50} (OECD Test Guideline 201 or equivalent). These species are considered as surrogate for all aquatic organisms and data on other species such as Lemna may also be considered if the test methodology is suitable.

Bioaccumulation potential

7. The potential for bioaccumulation would normally be determined by using the octanol/water partition coefficient, usually reported as a log Kow determined by OECD Test Guideline 107 or 117. While this represents a potential to bioaccumulate, an experimentally determined Bioconcentration Factor (BCF) provides a better measure and should be used in preference when available. A BCF should be determined according to OECD Test Guideline 305.

Rapid degradability

8. Environmental degradation may be biotic or abiotic (e.g. hydrolysis) and the criteria used reflect this fact (Annex I). Ready biodegradation can most easily be defined using the OECD biodegradability tests OECD Test Guideline 301 (A - F). A pass level in these tests can be considered as indicative of rapid degradation in most environments. These are freshwater tests and thus the use of the results from OECD Test Guideline 306 which is more suitable for marine environments has also been included. Where such data are not available, a BOD(5 days)/COD ratio >0.5 is considered as indicative of rapid degradation.

9. Abiotic degradation such as hydrolysis, primary degradation, both abiotic and biotic, degradation in non-aquatic media and proven rapid degradation in the environment may all be considered in defining rapid degradability. Special guidance on data interpretation will be provided in the Guidance Document.

Chronic toxicity

10. Chronic toxicity data are less available than acute data and the range of testing procedures less standardised. Data generated according to the OECD Test Guidelines 210 (Fish Early Life Stage), 202 Part 2 or 211 (Daphnia Reproduction) and 201 (Algal Growth Inhibition) can be accepted. Other validated and internationally accepted tests could also be used. The NOECs or other equivalent L(E)Cx should be used.

CLASSIFICATION CATEGORIES AND CRITERIA

11. Substances classified under the following criteria will be categorised as ‘hazardous to the aquatic environment’. These criteria describe in detail the classification categories detailed diagramatically in Annex 2.

Acute toxicity

Class: Acute I

Acute toxicity:

96 hr LC_{50} (for fish) ≤1 mg/L and/or

48 hr EC_{50} (for crustacea) ≤1 mg/L and/or

72 or 96hr ErC_{50} (for algae or other aquatic plants) ≤1 mg/L.

Class: Acute I may be subdivided for some regulatory systems to include a lower band at $L(E)C_{50}$ ≤0.1 mg/L.

Class: Acute II

Acute toxicity:

96 hr LC_{50} (for fish) >1 - ≤10 mg/L and/or

48 hr EC_{50} (for crustacea) >1 - ≤10 mg/L and/or

72 or 96hr ErC_{50} (for algae or other aquatic plants) >1 - ≤10 mg/L.

Class: Acute III

Acute toxicity:

96 hr LC_{50} (for fish) >10 - ≤100 mg/L and/or

48 hr EC_{50} (for crustacea) >10 - ≤100 mg/L and/or

72 or 96hr ErC_{50} (for algae or other aquatic plants) >10 - ≤100 mg/L.

Some regulatory systems may extend this range beyond an $L(E)C_{50}$ of 100 mg/L through the introduction of another class.

Chronic toxicity

Class: Chronic I

Acute toxicity:

96 hr LC_{50} (for fish) ≤1 mg/L and/or

48 hr EC_{50} (for crustacea) ≤1 mg/L and/or

72 or 96hr ErC_{50} (for algae or other aquatic plants) ≤1 mg/L

and the substance is not rapidly degradable and/or the log Kow ≥ 4 (unless the experimentally determined BCF <500).

Class: Chronic II

Acute toxicity

96 hr LC_{50} (for fish) >1 to ≤10 mg/L and/or

48 hr EC_{50} (for crustacea) >1 to ≤10 mg/L and/or

72 or 96hr ErC_{50} (for algae or other aquatic plants) >1 to ≤10 mg/L

and the substance is not rapidly degradable and/or the log Kow ≥4 (unless the experimentally determined BCF <500), unless the chronic toxicity NOECs are > 1 mg/L.

Class: Chronic III

Acute toxicity:

96 hr LC_{50} (for fish) >10 to ≤100 mg/L and/or

48 hr EC_{50} (for crustacea) >10 to ≤100 mg/L and/or

72 or 96hr ErC_{50} (for algae or other aquatic plants) >10 to ≤100 mg/L

and the substance is not rapidly degradable and/or the log Kow ≥4 (unless the experimentally determined BCF <500) unless the chronic toxicity NOECs are >1 mg/L.

Class: Chronic IV

Poorly soluble substances for which no acute toxicity is recorded at levels up to the water solubility, and which are not rapidly degradable and have a log Kow ≥ 4, indicating a potential to bioaccumulate, will be classified in this class unless other scientific evidence exists showing classification to be unnecessary. Such evidence would include an experimentally determined BCF <500, or a chronic toxicity NOECs >1 mg/L, or evidence of rapid degradation in the environment.

RATIONALE FOR THE SYSTEM

12. The system for classification recognises that the core intrinsic hazard to aquatic organisms is represented by both the acute and chronic toxicity of a substance, the relative importance of which is determined by the specific regulatory system in operation. Distinction can be made between the acute hazard and the chronic hazard and therefore separate hazard classes are defined for both properties representing a gradation in the level of hazard identified. The lowest of the available toxicity values will normally be used to define the appropriate hazard class(es). There may be circumstances, however, when a weight of evidence approach may be used. Acute toxicity data are the most readily available and the tests used are the most standardised. For that reason, these data form the core of the classification system.

13. Acute toxicity represents a key property in defining the hazard where transport of large quantities of a substance may give rise to short-term dangers arising from accidents or major spillages. Hazards classes up to $L(E)C_{50}$ values of 100 mg/L are thus defined although classes up to 1000 mg/L may be used in certain regulatory frameworks. The Acute: Class I may be further sub-divided to include an additional class for acute toxicity $L(E)C_{50}$ ≤0.1 mg/L in certain regulatory systems such as that defined by MARPOL 73/78 Annex II. It is anticipated that their use would be restricted to regulatory systems concerning bulk transport.

14. For packaged substances it is considered that the principal hazard is defined by chronic toxicity, although acute toxicity at $L(E)C_{50}$ levels ≤1 mg/L are also considered hazardous. Levels of substances up to 1 mg/L are considered as possible in the aquatic environment following normal use and disposal. At toxicity levels above this, it is considered that the short-term toxicity itself does not describe the principle hazard, which arises from low concentrations causing effects over a longer time scale. Thus, a number of hazard classes are defined which are based on levels of chronic aquatic toxicity. Chronic toxicity data are not available for many substances, however, and it is necessary to use the available data on acute toxicity to estimate this property. The intrinsic properties of a lack of rapid degradability and/or a potential to bioconcentrate in combination with acute toxicity may be used to assign a substance to a chronic hazard class. Where chronic toxicity is available showing NOECs >1 mg/L, this would indicate that no classification in a chronic hazard class would be necessary. Equally, for substances with an $L(E)C_{50}$ >100 mg/L, the toxicity is considered as insufficient to warrant classification in most regulatory systems.

15. While the current system will continue to rely on the use of acute toxicity data in combination with a lack of rapid degradation and/or a potential to bioaccumulate as the basis for classification for assigning a chronic hazard class, it is recognised that actual chronic toxicity data would form a better basis for classification where these data are available. It is thus the intention that the scheme should be further developed to accommodate such data. It is anticipated that in such a further development, the available chronic toxicity data would be used to classify in the chronic hazard in preference to that derived from their acute toxicity in combination with a lack of rapid degradation and/or a potential to bioaccumulate.

16. Recognition is given to the classification goals of MARPOL 73/78 Annex II which covers the transport of bulk quantities in ships tanks, which are aimed at regulating operational discharges from ships and assigning of suitable ship types. They go beyond that of protecting aquatic ecosystems, although that clearly is included. Additional hazard classes may thus be used which take account of factors such as physico-chemical properties and mammalian toxicity.

EXPLANATORY NOTES

17. The organisms fish, crustacea and algae are tested as surrogate species covering a range of trophic levels and taxa, and the test methods are highly standardised. Data on other organisms may also be considered, however, provided they represent equivalent species and test endpoints. The algal growth inhibition test is a chronic test but the EC_{50} is treated as an acute value for classification purposes. This EC_{50} should normally be based on growth rate inhibition. If only the EC_{50} based on reduction in biomass is available, or it is not indicated which EC_{50} is reported, this value may be used in the same way.

18. Aquatic toxicity testing by its nature, involves the dissolution of the substance under test in the water media used and the maintenance of a stable bioavailable exposure concentration over the course of the test. Some substances are difficult to test under standard procedures and thus special guidance will be developed on data interpretation for these substances and how the data should be used when applying the classification criteria.

19. It is the bioaccumulation of substances within the aquatic organisms that can give rise to toxic effects over longer time scales even when actual water concentrations are low. The potential to bioaccumulate is determined by the partitioning between n-octanol and water. The relationship between the partition coefficient of an organic substance and its bioconcentration as measured by the BCF in fish has considerable scientific literature support. Using a cut-off value of log P(o/w) $\geq$ 4 is intended to identify only those substances with a real potential to bioconcentrate. In recognition that the log P(o/w) is only an imperfect surrogate for a measured BCF, such a measured value would always take precedence. A BCF in fish of <500 is considered as indicative of a low level of bioconcentration.

20. Substances that rapidly degrade can be quickly removed from the environment. While effects can occur, particularly in the event of a spillage or accident, they will be localised and of short duration. The absence of rapid degradation in the environment can mean that a substance in the water has the potential to exert toxicity over a wide temporal and spatial scale. One way of demonstrating rapid degradation utilises the biodegradation screening tests designed to determine whether a substance is `readily biodegradable'. Thus a substance which passes this screening test is one that is likely to biodegrade `rapidly' in the aquatic environment, and is thus unlikely to be persistent. However, a fail in the screening test does not necessarily mean that the substance will not degrade rapidly in the environment. Thus a further criterion was added which would allow the use of data to show that the substance did actually degrade biotically or abiotically in the aquatic environment by >70% in 28 days. Thus, if degradation could be demonstrated under environmentally realistic conditions, then the definition of `rapid degradability' would have been met. Many

degradation data are available in the form of degradation half-lives and these can also be used in defining rapid degradation. Details regarding the interpretation of these data will be further elaborated in the Guidance Document. Some tests measure the ultimate biodegradation of the substance, i.e. full mineralisation is achieved. Primary biodegradation would not normally qualify in the assessment of rapid degradability unless it can be demonstrated that the degradation products do not fulfil the criteria for classification as hazardous to the aquatic environment.

21. It must be recognised that environmental degradation may be biotic or abiotic (e.g. hydrolysis) and the criteria used reflect this fact. Equally, it must be recognised that failing the ready biodegradability criteria in the OECD tests does not mean that the substance will not be degraded rapidly in the real environment. Thus where such rapid degradation can be shown, the substance should be considered as rapidly degradable. Hydrolysis can be considered if the hydrolysis products do not fulfil the criteria for classification as hazardous to the aquatic environment. A specific definition of rapid degradability is included as Annex 1. Other evidence of rapid degradation in the environment may also be considered and may be of particular importance where the substances are inhibitory to microbial activity at the concentration levels used in standard testing. The range of available data and guidance on its interpretation will be provided in the Guidance Document.

22. For inorganic compounds and metals, the concept of degradability as applied to organic compounds has limited or no meaning. Rather the substance may be transformed by normal environmental processes to either increase or decrease the bioavailability of the toxic species. Equally the use of bioaccumulation data should be treated with care. Specific guidance will be provided on how these data for such materials may be used in meeting the requirements of the classification criteria.

23. Poorly soluble inorganic compounds and metals may be acutely or chronically toxic in the aquatic environment depending on the intrinsic toxicity of the bioavailable inorganic species and the rate and amount of this species which may enter solution. A protocol for testing these poorly soluble materials is being developed and will be covered further in the special guidance.

24. The system also introduces as `safety net' classification (Class: Chronic IV) for use when the data available does not allow classification under the formal criteria but there are nevertheless some grounds for concern. The precise criteria are not defined with one exception. For poorly water soluble organic substances for which no toxicity has been demonstrated, classification can occur if the substance is both not rapidly degraded and has a potential to bioaccumulate. It is considered that for such poorly soluble substances, the toxicity may not have been adequately assessed in the short-term test due to the low exposure levels and potentially slow uptake into the organism. The need for this classification can be negated by demonstrating the absence of long-term effects, i.e. a long-term NOECs > water solubility or 1 mg/L, or rapid degradation in the environment.

25. While experimentally derived test data are preferred, where no experimental data are available, validated Quantitative Structure Activity Relationships (QSARs) for aquatic toxicity and log Kow may be used in the classification process. Such validated QSARs may be used without modification to the agreed criteria, if restricted to chemicals for which their mode of action and applicability are well characterised. Validity may be judged according to the criteria established within the USEPA/EU/Japan Collaborative Project. Reliable calculated toxicity and log Kow values should be valuable in the safety net context. QSARs for predicting ready biodegradation are not yet sufficiently accurate to predict rapid degradation.

ANNEX 1: RAPID DEGRADABILITY

Substances are considered rapidly degradable in the environment if the following criteria hold true:

a) if in 28-day ready biodegradation studies, the following levels of degradation are achieved;

- tests based on dissolved organic carbon: 70%

- tests based on oxygen depletion or carbon dioxide generation: 60% of theoretical maxima

These levels of biodegradation must be achieved within 10 days of the start of degradation which point is taken as the time when 10% of the substance has been degraded.

or

b) if, in those cases where only BOD and COD data are available, when the ratio of BOD5/COD is ≥0.5

or

c) if other convincing scientific evidence is available to demonstrate that the substance can be degraded (biotically and/or abiotically) in the aquatic environment to a level >70% within a 28 day period.

ANNEX 2: Classification Scheme for Substances Hazardous to the Aquatic Environment

Toxicity		Degradability (note 3)	Bioaccumulation (note 4)	Classification categories	
Acute (note 1)	Chronic (note 2)			Acute	Chronic
Box 1 value ≤ 1.00		**Box 5**	**Box 6**	**Class: Acute I** **Box 1**	**Class: Chronic I** **Boxes 1+5+6** **Boxes 1+5** **Boxes 1+6**
Box 2 1.00 < value ≤ 10.0		lack of rapid degradability	BCF ≥ 500 or, if absent log Kow ≥ 4	**Class: Acute II** **Box 2**	**Class: Chronic II** **Boxes 2+5+6** **Boxes 2+5** **Boxes 2+6** **Unless Box 7**
Box 3 10.0 < value ≤ 100				**Class: Acute III** **Box 3**	**Class: Chronic III** **Boxes 3+5+6** **Boxes 3+5** **Boxes 3+6** **Unless Box 7**
Box 4 No acute toxicity **(note 5)**	**Box 7** value > 1.00				**Class: Chronic IV** **Boxes 4+5+6** **Unless Box 7**

Notes to the table:

Note 1a. Acute toxicity band based on L(E)C-50 values in mg/L for fish, crustacea and/or algae or other aquatic plants (or QSAR estimation if no experimental data)

Note 1b Where the algal toxicity ErC-50 [= EC-50 (growth rate)] falls more than 100 times below the next most sensitive species and results in a classification based solely on this effect, consideration should be given to whether this toxicity is representative of the toxicity to aquatic plants. Where it can be shown that this is not the case, professional judgement should be used in deciding if classification should be applied. Classification should be based on the ErC-50. In circumstances where the basis of the EC-50 is not specified and no ErC-50 is recorded, classification should be based on the lowest EC-50 available.

Note 2a. Chronic toxicity band based on NOEC values in mg/L for fish or crustacea or other recognised measures for long-term toxicity.

Note 2b. It is the intention that the system be further developed to include chronic toxicity data.

Note 3. Lack of rapid degradability is based on either a lack of Ready Biodegradability or other evidence of lack of rapid degradation.

Note 4. Potential to bioaccumulate, based on an experimentally derived BCF ≥ 500 or, if absent, a log Kow ≥ 4 provided log Kow is an appropriate descriptor for the bioaccumulation potential of the substance. Measured log Kow values take precedence over estimated values and measured BCF values take precedence over log Kow values.

Note 5. "No acute toxicity" is taken to mean that the L(E)C-50 is above the water solubility. Also for poorly soluble substances, (w.s. < 1.00 mg/L), where there is evidence that the acute test would not have provided a true measure of the intrinsic toxicity.

HARMONIZED SYSTEM FOR THE CLASSIFICATION OF CHEMICAL MIXTURES

(to be inserted here after approval by the OECD Advisory Group on Harmonization of Classification and Labelling and subsequent endorsement by the Joint Meeting)

Purpose, basis and applicability

Definitions

Classification categories and criteria

Rationale for the proposed System

Explanatory notes

Literature

Background information

SCHEMATIC PRESENTATION OF THE INTEGRATED HAZARD CLASSIFICATION SYSTEM FOR HUMAN HEALTH AND THE ENVIRONMENT

For the convenience and comparison of the various endpoints, the scheme and criteria for classifying each toxic end-point are presented in the following diagram. The criteria have been drastically abridged and the end-point chapters must be consulted for the specific details to avoid misunderstanding.

ENDPOINT	HAZARD CLASSES AND CRITERIA

ACUTE TOXICITY	Class 1	Class 2	Class 3	Class 4	Class 5
Oral (mg/kg)	5	50	300	2 000	5 000 Criteria: • Indication of significant effect in human • Any mortality at Class 4 • Significant clinical signs at Class 4 • Indications from other studies
Dermal (mg/kg)	50	200	1 000	2 000	
Inhalation note 1 gas (ppm)	100	500	2 500	5 000	
vapour (mg/L) note 2,3	0.5	2.0	10	20	
dust/mists (mg/L/4 hrs) note 4	0.05	0.5	1.0	5	

Note 1: Inhalation cut-off values are based on 4 hour testing exposures. Conversion of existing inhalation toxicity data which has been generated according to 1 hour exposures should be by dividing by a factor of 2 for gases and vapours and 4 for dusts and mists.

Note 2: Saturated vapour concentration may be used as an additional element to provide for specific health and safety.

Note 3: For some chemicals the test atmosphere will not just be a vapour but will consist of a mixture of liquid and vapour phases. For other chemicals the test atmosphere may consist of a vapour which is near the gaseous phase. In these latter cases, classification should be based on ppm as follows: Class 1 (100 ppm), Class 2 (500 ppm), Class 3 (2500 ppm), Class 4 (5000 ppm).

Note 4: The values for dusts and mists should be reviewed to adapt to any future changes to OECD Test Guidelines with respect to technical limitation in generating, maintaining and measuring dust and mist concentrations in respirable form.

ENDPOINT	HAZARD CLASSES AND CRITERIA

DERMAL IRRITATION/ CORROSION	Class 1			Class 2:	Class 3:
	Destruction of dermal tissue: visible necrosis in at least one animal			- Reversible adverse effects in dermal tissue	- Reversible adverse effects in dermal tissue
	Subclass 1A Exposure < 3 minutes Observation < 1 hour	Subclass 1B Exposure < 1 hour Observation < 14 days	Subclass 1C Exposure < 4 hours Observation < 14 days	- Mean Draize score in 2 of 3 animals: 2.3 ≤erythema/eschar/ edema < 4.0, or - persistent inflammation	- Mean Draize score in 2 of 3 animals: 1.5 ≤ erythema/ eschar/ edema < 2.3

EYE IRRITATION/ CORROSION	Class 1 - Irreversible damage to cornea, iris, conjunctiva 21 days after exposure in at least one animal - mean Draize score in 2 of 3 animals: corneal opacity ≥ 3, iritis >1.5	Class 2 - reversible adverse effects on cornea, iris, conjuctiva - mean Draize score in 2 of 3 animals: corneal opacity: ≥1, iritis: ≥1, redness ≥ 2, chemosis: ≥2
		Subclass 2A: reversible in 21 days / Subclass 2B: reversible in 7 days

RESPIRATORY SENSITISATION	Class 1: - evidence of specific respiratory hypersensitivity, or - positive results from animal test

DERMAL SENSITISATION	Class 1: - evidence in humans of sensitisation by skin contact, or - positive results from animal tests

ENDPOINT	HAZARD CLASSES AND CRITERIA			
GERM CELL MUTAGENICITY	Class 1 known to produce heritable mutations in human germ cells		Class 2:	
	Subclass 1A positive evidence from epidemiological studies	Subclass 1B positive results in: - *in vivo* heritable germ cell tests in mammals - human germ cell tests - *in vivo* somatic mutagenicity tests, combined with some evidence of germ cell mutagenicity	- may induce heritable mutations in human germ cells - positive evidence from tests in mammals and somatic cell tests - *in vivo* somatic genotoxicity supported by *in vitro* mutagenicity	
CARCINOGENICITY	Class 1: Known or presumed carcinogen		Class 2: - suspected carcinogen - limited evidence of human or animal carcinogenicity	
	Subclass 1A: known human carcinogen based on human evidence	Subclass 1B: presumed human carcinogen based on demonstrated animal carcinogenicity		
REPRODUCTIVE TOXICITY	Class 1: known or presumed human reproductive or developmental toxicant		Class 2: suspected human reproductive toxicant	Additional Class effects on or via lactation
	Class 1A: known	Class 1B: presumed		

ENDPOINT	HAZARD CLASSES AND CRITERIA

TARGET ORGAN ORIENTED TOXICITY	(to come)

AQUATIC TOXICITY	Acute Class 1: acute toxicity ≤ 1.00mg/L	Acute Class 2: acute toxicity > 1.00 but ≤ 10.0mg/L	Acute Class 3: acute toxicity > 10.0 but ≤ 100mg/L	
	Chronic Class 1: acute toxicity ≤ 1.00mg/L and lack of rapid degradability and log Kow ≥ 4 unless BCF < 500	Chronic Class 2: acute toxicity > 1.00 but ≤ 10.0mg/L and lack of rapid degradability and log Kow ≥ 4 unless BCF < 500 and unless chronic toxicity > 1 mg/L	Chronic Class 3: acute toxicity > 10.0 but ≤ 100mg/L and lack of rapid degradability and log Kow ≥ 4 unless BCF < 500 and unless chronic toxicity > 1mg/L	Chronic Class 4 acute toxicity > 100 mg/L and lack of rapid degradability and log Kow ≥ 4 unless BCF < 500 and unless chronic toxicity > 1mg/L

MIXTURES	(to come)